DOCUMENT RESUME

ED 287 213 EA 019 728

AUTHOR	Weber, James R.
TITLE	Teacher Evaluation as a Strategy for Improving Instruction. Synthesis of Literature.
INSTITUTION	North Central Regional Educational Lab., Elmhurst, IL.
SPONS AGENCY	Office of Educational Research and Improvement (ED), Washington, DC.
PUB DATE	Jul 87
CONTRACT	400-86-0004
NOTE	72p.; Prepared by the ERIC Clearinghouse on Educational Management under contract to NCREL. For other documents in the same series, see EA 019 726-731.
AVAILABLE FROM	Publication Sales, ERIC Clearinghouse on Educational Management, University of Oregon, 1787 Agate St., Eugene, OR 97403 ($10.00); North Central Regional Educational Laboratory, 295 Emroy Ave., Elmhurst, IL 60126 ($10.00).
PUB TYPE	Information Analyses - ERIC Information Analysis Products (071) -- Reports - Evaluative/Feasibility (142)
EDRS PRICE	MF01/PC03 Plus Postage.
DESCRIPTORS	Accountability; Classroom Observation Techniques; Elementary Secondary Education; *Evaluation Methods; Feedback; *Instructional Improvement; Lesson Observation Criteria; Teacher Administrator Relationship; Teacher Attitudes; Teacher Effectiveness; *Teacher Evaluation; *Teacher Improvement; *Teacher Supervision
IDENTIFIERS	ERIC Clearinghouse on Educational Management

ABSTRACT

This review surveys major currents of thought and practice in teacher evaluation. Citing recent state-mandated teacher evaluation policies and procedures, several compelling questions of accuracy, fairness, and utility are raised. In response to these questions, the document first focuses on the distinction between formative and summative evaluation, the relation of the latter to minimum standards and legal mandates, and the correlation of methods with purposes. Alternatives developed since the 1960s include goal-setting models such as the performance-objectives approach, outcome-based models, and clinical supervision models. In the next section, the separate problems of the two main participants in the teacher evaluation process--evaluators and teachers--are analyzed. The evaluator's main concerns are the separation of summative and formative tasks, the need for expertise, and the relationship with the teacher, while teachers need to be involved in developing evaluation criteria, and they need to feel that the criteria by which they are evaluated are sound and relevant to their teaching. The fourth section discusses the three stages of teacher evaluation--preobservation conferences, the observation itself, and the postobservation conference--and touches on other sources of data besides observation: parent evaluations, peer observation, teaching materials, student evaluations, and self-evaluations. The conclusion addresses four key issues: (1) coexistence of teacher development and accountability; (2) supervision versus evaluation; (3) utility of evaluation in improving teaching; and (4) the most productive, least time-wasting approaches to observation. Appended is a syllabus of Thomas McGreal's training program for staff and supervisors. (TE)

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

_____

_____

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Teacher Evaluation
as a Strategy for Improving Instruction

Synthesis of Literature

James R. Weber

July 1987

Prepared for the
North Central Regional Educational Laboratory

by the
ERIC Clearinghouse on Educational Management
University of Oregon

Foreword

The ERIC Clearinghouse on Educational Management at the University of Oregon and the North Central Regional Laboratory at Elmhurst, Illinois, are pleased to offer this publication, part of a series of syntheses papers and annotated bibliographies on themes related to instructional leadership and school improvement. The Clearinghouse wrote and edited the materials under a subcontract for the North Central Laboratory. Both agencies are now making the publications available to their respective clienteles.

The titles of all the publications in this series are as follows:

Synthesis Papers
Instructional Leadership: A Composite Working Model
Teacher Evaluation as a Strategy for Improving Instruction
From Isolation to Collaboration: Improving the Work Environment of Teaching

Annotated Bibliographies
Models of Instructional Leadership
Teacher Evaluation
The Social and Organizational Context of Teaching

The author of this publication, James R. Weber, is a research analyst and writer for the ERIC Clearinghouse on Educational Management.

Jane H. Arends
Executive Director
North Central Regional
Educational Laboratory

Philip K. Piele
Director
ERIC Clearinghouse on
Educational Management

About ERIC

The Educational Resources Information Center (ERIC) is a national information system operated by the National Institute of Education. ERIC serves the educational community by disseminating educational research results and other resource information that can be used in developing more effective educational programs.

The ERIC Clearinghouse on Educational Management, one of several clearinghouses in the system, was established at the University of Oregon in 1966. The Clearinghouse and its companion units process research results and journal articles for announcement in ERIC's index and abstract bulletins.

Research reports are announced in *Resources in Education (RIE)*, available in many libraries and by subscription for $51.00 a year from the United States Government Printing Office, Washington, D.C. 20402.

Most of the documents listed in RIE can be purchased through the ERIC Document Reproduction Service, operated by Computer Microfilm International Corporation.

Journal articles are announced in *Current Index to Journals in Education*. *CIJE* is also available in many libraries and can be ordered for $150.00 a year from Oryx Press, 2214 North Central at Encanto, Phoenix, Arizona 85004. Semiannual cumulations can be ordered separately.

Besides processing documents and journal articles, the Clearinghouse prepares bibliographies, literature reviews, monographs, and other interpretive research studies on topics in its educational area.

Contents

Introduction 1

The Growing Interest in Teacher Evaluation 1
Compelling Questions 2
Dimensions of Evaluation 3
Purpose of This Synthesis 3

1. The Context of Teacher Evaluation 5
Summative versus Formative Evaluation 5
Common-Law Evaluation 9
Alternative Models 11
Conclusion 18

2. The Evaluators: Who They Are and What They Do 22
Separation of Summative and Formative Tasks 22
Expertise Needed by Evaluators 24
Relationship between Evaluator and Teacher 28

3. The Teachers: Concerns and Participation 33
Teacher Participation and Program Success 35
Other Human-Factor Suggestions 38
Conclusion 39

4. Appropriate Data and Effective Feedback 41
Preobservation Conferences 41
Observations and Data 43
Other Sources of Data 48
The Dynamics of Feedback 52

Conclusion 56

Appendix 58

Bibliography 61

Introduction

The history of teacher evaluation in the public schools of the United States has been marked by a tension between including teachers' input and applying standards from administrative criteria. From the last decade of the nineteenth century, teacher evaluation has been increasingly "humanized" by including more concerns for the development of teachers' skills.

Before the turn of the century, teacher "inspection" was the evaluation method most frequently practiced. Administrators, who did not need to be trained in teaching or observing, observed teachers for their conformity to district expectations. These expectations could be personal as well as professional. Evaluations might focus on critiques of student behavior, for instance, or on a teacher's personality--including out-of-class activities.

The emphasis then shifted to the efficiency of teaching and the "scientific management" of students and school personnel. After this interest in efficiency and economy of effort, however, administrators began to see the need to cooperate with teachers in evaluations. Researchers, too, began to isolate significant teaching behaviors, warranting the belief that good teaching can be developed with adequate attention and effort. By the post-World War II period, cooperation between supervisors and teachers was an assumption in the research, if not in the majority of schools (Clara Peterson 1982). Through the influence of clinical supervision approaches, concern for mutual effort and reciprocity are features of nearly all new models of supervision.

On the other hand, as evaluation turned more democratic in theory (if not in the practice of most schools), it has been matched by a growing public pressure for teacher accountability. The result has been numerous programs that combine the historical gains in development-centered evaluations with accountability strategies aimed at ensuring minimum standards and encouraging maximum effort.

The Growing Interest in Teacher Evaluation

In the effort to improve teaching, a great deal of energy has recently been directed at improving teacher evaluations. At the policy-making level, states and school districts have been initiating programs to accelerate schools' procedures for dismissing incompetent teachers or improving competent teachers. For example, in 1985 Kansas began a statewide, legislatively approved internship program for teachers. Local committees, consisting of administrators and senior teachers, will assess,

assist, and support all first-year teachers. That same year sixteen Arizona school districts submitted plans for career ladders for their teachers. Under legislation passed by that state's legislature in 1984, districts' career ladders must include objective performance criteria for advancement and more than one measure of performance, one of which must measure student achievement (Ross and Solomon 1985).

Compelling Questions

The career-ladder program begun by the Tennessee Legislature in 1984 sets out three basic levels, time lines, and evaluation criteria and procedures. Teachers are measured by a combination of skills tests, observations, and evaluations by students, colleagues, and principals. In Georgia, a program of assessment for first-year teachers, begun in 1980, blossomed into a career-ladder program for all teachers (Ross and Solomon). As of 1983, twenty-six states had teacher-evaluation laws, 80 percent of which were enacted since 1971 (Stiggins and Bridgeford 1985).

Although these programs seem to be steps in the right direction, they also raise compelling questions of accuracy, fairness, and utility: Can state-mandated evaluation processes ensure that the gains in the humanization of teacher evaluation of the last century will be continued? How can teacher development strategies coexist with accountability strategies? Can the same people who decide teachers' career placement also oversee their professional development? How useful are evaluation programs for improving teaching? What specific approaches to classroom observation are the most productive and least wasteful?

Furthermore, there seems to be a disturbing discrepancy between administrators' and teachers' views about the supervisory services being provided for teachers. In a survey of teachers and administrators conducted by the Association for Supervision and Curriculum Development, supervisors and administrators consistently rated the quality of their instructional supervision higher than teachers did (Cawelti and Reavis 1980).

In general, although evaluation procedures are becoming more systematic, the help they offer to teachers for improving their teaching varies widely from program to program. As Richard Stiggins and Nancy Bridgeford have noted, issues of money and time may prevent districts from helping teachers to career improvement:

> Teachers want, at the very least, an evaluation system that provides accurate information on classroom needs, opportunity to acquire and master new learning approaches, and collegial support when instigating needed changes. These activities demand more time, instructional involvement, and more thorough assessment

than many principals seem to find manageable. As a consequence, [teaching] practices become more formalized, remaining basically unchanged.

Dimensions of Evaluation

What can we do to untangle the numerous threads of needs, personal interests, and varied experiences that make up current discussions of teacher evaluations? Luckily, some research-based theories can help us divide and discuss the elements that comprise evaluations. One insightful approach, conceived by Daniel Duke and Richard Stiggins (1986), divides the evaluation process into five attributes that can be considered separately as dimensions of evaluation:

the teachers

the evaluators

the performance data

the feedback

the context of the evaluation

These attributes suggest that evaluations are social and personal, objective and individualistic, overt and subtle, immediate and repercussive. As the following chapters show, teacher evaluations may be one of the most potent tools for improvement--or for stagnation--available to those who seek to influence schools.

Purpose of this Synthesis

This study is designed to be a mirror of the issues surrounding teacher evaluation as they are found in the research of the late 1980s. It is not intended as an original approach to the practice of teacher evaluation, nor an exhaustive compilation of a field with exhaustingly large and vital potential. Teacher evaluation is (as these pages reflect) a complex social, psychological, and managerial challenge. Instead, this is a state-of-the-art survey, reflecting major currents of thought and practice in evaluation. It is hoped that practitioners and those responsible for planning evaluation programs can use this study for organizing their thinking, be stimulated to read in more depth about the ideas presented here, or create their own new directions to overcome the present limitations in evaluation programs.

The paper begins by reviewing the common practices of teacher evaluation and the alternative approaches developed since the 1960s. Then, the separate problems of the main participants in

the process are analyzed, beginning with evaluators and then proceeding to teachers. Having discussed the human factors in evaluation--ultimately, the most important aspect--attention turns to the mechanics of assessment and the recycling of results to stimulate teacher development.

Chapter 1

The Context of Teacher Evaluation

Nearly everyone agrees that the ultimate aim of teacher evaluations is to create competent, effective teachers who will improve student performance. But the road toward this goal is strewn with controversies. Teacher evaluation has become an issue of conflicting social interests that interfere with attempts to build practical evaluation programs for schools.

Many policy makers--claiming they represent the public's will--have decided that the most direct way to improve student achievement is to emphasize teachers' accountability, using tests and other means to weed out the ineffective and incompetent teachers. Teachers, on the other hand, prefer evaluation systems that are meant to improve teaching. They want evaluations that preserve the autonomy and rights of teachers and that take into account the complexity of the teaching art. Principals and district administrators, caught between the political pressures of public and teachers, also have their interests in evaluation, one of which is maintaining a stable organization with good morale and few unnecessary staff problems.

Summative versus Formative Evaluation

The kinds of teacher evaluations used also reflect this division of interests. Accountability advocates prefer a *summative* evaluation model, rating teachers against a fixed scale of standards and then comparing their performances against their colleagues'. Summative models may be convenient for ranking teachers according to merit and eliminating incompetent teachers. Such models appeal to advocates of merit pay and master teacher plans.

Formative evaluations concentrate on pinpointing teachers' weaknesses and strengths toward making them better teachers. Most formative models are "feedback" models, with multiple evaluations spread over an extended period. Coaching may be provided for teachers, and formative models can be connected with staff development activities. Unfortunately, as Rand Corporation researchers found (Wise and others 1985), links to staff development rarely exist.

Minimum Standards and Legal Mandates

Most districts have adopted summative models so they can better defend teacher dismissal procedures in court. Courts have required that districts have policies setting out minimum

acceptable standards for their teachers, provide information to the staff about standards for competence, and document each teacher's performance, showing clearly how it satisfies or violates the minimum standards (Beckman 1981). Consequently, many schools have simply extended their old evaluation procedures (nearly always summative) into more structured, legally mandated forms. Whereas principals used to observe teachers perhaps once a year and answer some broad questions about the teacher's performance, now they are encouraged to fill out carefully constructed questionnaires about the teacher's performance, suitable for presentation in court. In such teacher evaluations, schools have a ready-made due-process safeguard.

In sum, the courts (and the realities of collective bargaining) have forced schools to approach evaluation as a rule-based process instead of a professional procedure. In many schools, the minimum standards for competency set down by the state or local district have become synonymous with evaluations.

Of course, such safeguards do not necessarily contribute to improving teaching because they are primarily *minimum* standards. Indeed, to the extent they inhibit teachers from developing more-than-minimum performances, their effect on schools will be negative. When a climate of punitive evaluation pervades a school, teachers will resist evaluation or treat it as a threat. Another weakness of accountability-directed systems is that they have a lasting effect on only a few teachers--those who are not competent. Thus, in the quest to improve overall school quality, those systems will provide slow progress (Stiggins 1986).

Moreover, there is no general agreement on how to measure even "minimum" teacher competencies. Detecting incompetence requires reliable and generalizable measurements that may not always be possible given the present state of our knowledge about teaching and learning. Correcting weaknesses in teaching is possible, but this requires a formative rather than a summative approach to evaluations--and thus the commitment of districts to an ongoing process of evaluation.

Summative and Formative Systems Complement Each Other

In general, though a summative approach is a necessary component of evaluation systems, it is not sufficient. What the courts require (and many collective bargaining agreements are demanding) is a suitable formative evaluation process that accompanies the summative process. The formative model requires that administrators and teachers not simply go through the motions of improvement but manifest a real commitment, one in which the public, unions, courts, and teachers can believe. To be effective, then, summative and formative systems must go hand in hand.

In practice, evaluation and supervision serve a variety of practical purposes. Looking at the major uses to which summative and formative evaluations are put, we may find it very difficult to see how one system could be used effectively or accurately without the other. In a recent Rand Corporation study of teacher evaluation procedures, school administrators cited four purposes for evaluations: personnel decisions involving teacher placement and tenure; staff development, such as identifying areas for teacher inservice training; school improvement, focusing on upgrading the general level of instruction (as in overall instructional goals for schools or departments); and accountability, centering on meeting or exceeding district and state standards (Wise and Darling-Hammond 1984-85).

Some school districts claim to be meeting all these goals with a single evaluation system--that is, a single measurement instrument and a single supervision process. But can a single method serve all these purposes? If a major goal of an evaluation system is to eliminate incompetent teachers, can it also help all teachers improve? With the interest in merit pay and master teachers, many districts appear to want evaluations that rank, monitor, and cull the chaff from the faculty.

The two types of systems differ in breadth of coverage (the summative systems reach many more teachers) and in depth (the formative systems expose teachers' plans and styles in considerably more detail). They differ in the way in which each recognizes good teaching: summative methods use a standardized approach; formative methods use a context-specific, individualized approach. They also differ in the kinds of evidence they gather about teachers' abilities. Accountability systems must protect everyone's due-process rights. Thus, the information considered important must be both consistent with preset criteria and legally admissible. "With accountability, legal requirements preclude the use of most of the valuable sources of information on performance" (Stiggins 1986). It is mistaken, then, to think that one purely summative or formative system can serve the purposes of growth, accountability, school improvement, and personnel decisions.

Multiple Purposes Require Multiple Methods and Data

If a school district wants to achieve multiple purposes with an evaluation, the district should consider using more than one method. Would a district want to promote a teacher to master status, for example, on the basis of only minimum standards? Undoubtedly not, but exclusively using accountability systems might lead to their doing so. Merit pay and master teacher programs both require rigorous evaluation methods, but they also may require different levels of effort. Because merit pay has visible consequences--creating *de facto* a pay differential among

faculty--evaluation for this purpose must be as rigorous and credible as that for dismissals. "A school district that intends to evaluate all teachers annually for merit pay decisions must commit substantial resources to evaluation" (Wise and Darling-Hammond 1984-85).

Both merit pay and master teacher programs, moreover, differ from termination-directed evaluation in the kinds of data useful to making decisions and in the nature of the evaluators. Evaluation for termination distinguishes between inadequate and minimally adequate teachers, whereas evaluation for excellence (as in the case of merit pay and master teachers) distinguishes between marginally excellent and highly competent teachers.

Table 1
Comparison of Summative and Formative Models

	Formative Evaluation
Rating Scales	uses flexible criteria; emphasize teaching context
Outcome	advises teacher on improvement
Evaluators	to be effective, must have teaching background, plus knowledge of each teacher's strategies
Time Demands	may require repeated visits, conferences, and analysis of teaching materials
Data Sources	relies on observations, teaching materials, student scores, plus information from teachers on intentions and perceptions (self-assessment, peer assessment), climate
Motivation for Teacher Improvement	relies on teachers' desire to improve
Primary Purpose	fosters professional development

Moreover, generalists (such as principals) can evaluate for minimal competence, but experts must judge excellence in subject areas and in matters of teacher improvement. Thus, although rigorous systems are required for purposes of reward or dismissal, there may be considerable differences in implementation and conditions for success.

The major points of difference between summative and formative models of teacher evaluation are summarized in Table 1.

Common-Law Evaluation

Many districts, however, attempt to meet multiple goals with an all-purpose evaluation system. Thomas McGreal (1983) has found certain features so common in evaluation systems that he calls them "common law evaluations"--districts have been married to them by simply living with them for so long. These systems give lip service to teacher improvement as their prime purpose, but then provide only for termination or tenure evaluations. Formative on the surface, they are summative in operation. Parts of these systems may serve some needs in particular districts, but they also provide the most negative image of teacher evaluation in current use. McGreal estimates that 65 percent of school districts in the United States use some form of the common-law method.

Common-law systems rely on simple definitions of evaluation and a minimum of processes, as this typical opening statement reveals:

> GENERAL STATEMENT:
> This district believes that each child has unique educational and socio-emotional needs that require quality instruction by all staff members. The district and its professional employees have a responsibility to see that the needs of the students are being met. One way to meet this responsibility is to have a teacher evaluation procedure that is designed to improve the quality of instruction. In order to be most effective, the procedure should involve both teachers and administrators throughout the process.
>
> PROCEDURES:
>
> (1) All nontenured staff will be evaluated by their principal at least three times during the school year. A professional evaluation form must be submitted after each evaluation. The final report must be on file no later than the end of the first week in March.
> (2) All tenured teachers will be evaluated by the principal or his or her designee at least once each school year. A professional evaluation report must be

submitted by April 15.
(3) A conference must be held with the staff member following each evaluation. The completed evaluation report must be reviewed with the staff member during the conference. Suggestions for improving areas marked fair or weak should be made along with plans for any follow-up visits. Both parties should then sign the report.
(4) Teachers have the option to write comments about any part of the evaluation in the appropriate space.
(McGreal 1983)

This preamble exemplifies several characteristics of common-law models. First is a high-supervisor/low-teacher involvement in the evaluation process, with the teacher being a relatively passive participant. The supervisor determines when visits will be scheduled, fills out the required forms, and conducts the post-evaluative conference. Second, evaluation is generally seen as synonymous with observations; little or no data other than classroom visits are used in evaluating. Third, procedures do not vary for tenured and nontenured teachers, though nontenured teachers are evaluated more often.

Fourth, the major purpose of evaluation is for summative judgments, usually for personnel decisions. There is no attention given to the other purposes identified by the Rand study: staff development, school improvement, and accountability. The evaluation tells teachers where they stand in relation to others instead of what they are doing and how they might improve. As with most summative evaluation strategies, there is a standardized set of traits on which teachers are measured.

Certainly, school districts have had good reason to be married to this sort of evaluation process for so long. It has great utility. It can be used economically where there are many teachers and few supervisors. (McGreal finds that whenever a supervisor is responsible for more than twenty teachers' evaluations annually, the common-law model is economical.) Generally, the requirements do not demand extensive supervision. This model also requires very little training for supervisors. It allows generalists--which is what most principals are--to apply standard criteria rather than special knowledge of subject areas.

Furthermore, districts that use the common-law model can appear to meet accountability demands while avoiding the sensitive areas that may be disruptive to staff. The straightforward, heavily supervisory method looks good from the outside: hard-working administrators are doing their jobs in ways that school boards and noneducators can understand. This model also has the advantage of being nonthreatening to teachers, who can sustain one or two evaluations per year without a threat to position or teaching style. As an evaluation tool, the common-law model is also reliable. That is, several evaluators can use the same

standardized criteria to reach the same conclusions.

So why are researchers and school reformers almost universally critical of the common-law model for evaluation? Although it may rate high in utility and relative reliability, it is not a *valid* way of judging teaching. Validity refers to the accuracy and comprehensiveness of the evaluation in the context of the classroom. Teacher evaluators using the common-law model may overlook the most valuable evidence of a teacher's competence. McGreal notes that as many as 70 percent of the criteria on common-law evaluations deal with administrative or personal concerns rather than with instructional performance; that is, the criteria may cover teacher-staff-parent relationships, professional participation, or recordkeeping.

Finally, the evaluators in a common-law method need not know anything about teaching in a particular area. .The model itself, it appears, works against validity in evaluations. In such a system, where utility and reliability are high but validity is low, the same mistakes in judgment can be made again and again. The lowest common denominators will prevail.

Alternative Models

The reason for the flawed validity of common-law models may be that they are normally used for their expediency and convenience rather than from a real commitment to teacher improvement. Although they may fill the district's need for summative evaluations, such models are not consciously designed to fulfill other purposes, stated or not. The Rand study found that administrators in seventeen of twenty districts had difficulty specifying the primary goals of their teacher evaluation systems (Wise and Darling-Hammond 1984-85). This response probably reflects the general perception that neither summative nor formative needs are being met by present methods.

Effective evaluation systems share the trait of being consciously designed to meet a particular district's needs: dismissal decisions, master teacher or merit pay programs, staff development strategies, individual teacher improvements, and so forth. In customizing an evaluation model, districts have generally chosen some variation or combination of four patterns: goal-setting models, product models, clinical supervision models, and naturalistic (or artistic) models. The first three are introduced here, with the fourth discussed in chapter 4.

Goal-Setting Models

Goal-setting models avoid problems of validity by involving the teacher in determining the criteria for evaluation. The

teacher actually begins the process by conducting a self-evaluation, noting those areas in which he or she feels weakest. The teacher then drafts a goal-setting "contract," after which teacher and evaluator meet to discuss the self-evaluation, contract, and steps needed to improve. The evaluator then confers periodically with the teacher to monitor progress toward the contracted goals. Finally, at the close of the agreed evaluation cycle, they examine the results of the effort and plan for future improvement goals. The high teacher-involvement keeps the criteria meaningful to teachers; the preconferences and postconferences introduce reliability into the evaluations as well.

Performance-objectives approach. One such program has been proposed by George B. Redfern (1980). This performance-objectives approach, as described by Redfern, arose as a reaction against schools evaluating teachers' personalities or other factors extraneous to directly measurable teaching criteria. The heart of the plan is the setting of objectives, forming an action plan, and then carrying out and monitoring the results:

> With this approach, particular areas or problems of performance are identified. For example, a teacher may indicate a desire to improve discipline in the classroom. This is a real problem that has a direct bearing upon effectiveness in teaching. The teacher and the principal discuss the matter. They may agree that this calls for a single objective. An understanding is reached as to the procedure that will be followed to accomplish improvement. Agreement is reached about the way success or the lack of it is to be determined. At the end of the year, the evaluator, in cooperation with the teacher, will make a judgment about progress made in attaining desired results (Redfern 1980).

The performance-objectives approach rests on how several essential features are arranged. Job duties must be specified, preferably by a detailed list of responsibilities. Job descriptions commonly used in personnel recruitment would leave too much to personal interpretation and almost inevitably lead to misunderstandings in the evaluation process. Objectives, then, can reflect some aspect of these detailed responsibilities. Rather than using generally stated objectives, participants should use behavioral objectives to facilitate mutual understanding and ease of documentation. Moreover, a single written form can contain both the performance objective and the action plan; the teacher and supervisor, then, can both understand what is to be done, the outcome desired, and the method of measurement used.

The assessment of results, despite the careful mutual planning throughout the process, might well lead to disagreements between teacher and evaluators over whether the objective has been

achieved. To anticipate this possibility, Redfern's program includes a structured teacher self-evaluation. A written summative report should also be included on the list of job responsibilities with which the cycle began.

Constraints and benefits. Like magnifying glasses that focus the sun's rays, goal-setting models limit and concentrate the energy of teachers and evaluators. They are obviously formative rather than summative models. A goal-setting model, then, is probably not suitable for ranking teachers. Also, much depends on the contract that the teacher draws up and the evaluator reviews. The contract must specify observable, measurable behaviors or outcomes and must identify the acceptable outcomes. It must further provide a date for accomplishing the goals. Useful goals may be hard to form: they must be realistic and yet challenging, attainable with existing resources, and consistent with departmental, school, and district goals.

Despite these drawbacks and constraints, the benefits of a goal-setting model are considerable. They focus attention on the professional growth of each teacher, rather than settling for a lowest common denominator. They also encourage a working relationship between teachers and supervisors--breaking down the barriers that have been described as a "private cold war." One obvious benefit of this relationship is the clarification of performance expectations, making the criteria unambiguous and personal.

Product Models

Product models assume that teachers can best be evaluated by measuring student achievement. If teachers can produce high student-competency in an area, then teacher competency must also be high in that area. Because much depends on being able to measure student achievement accurately, the nature of the tests used in product evaluation models is a primary issue. Generally, a time period is designated for the evaluation cycle, with a pretest (or a guess about expected performance) and a posttest administered to show any changes in student ability. Norm-referenced tests (measuring the student performance on a curve) may be used, as may criterion-reference tests (measuring performance according to a preset standard).

Distinguishing student achievement from competence. Perhaps the simplest and most inaccurate method of judging teacher performance is to compare the raw scores of students on standardized achievement tests. The trouble with this practice lies in its confusion of student *achievement* with student *competence*. Achievement can be defined as what an individual can do or knows as a consequence of instruction. This is certainly what we would need to measure for the purposes of teacher

evaluation. However, so-called "achievement tests" most often do not measure instruction-generated achievement, but measure instead the student's competence--a student's cumulative knowledge about a subject acquired through varied experience and (probably) more than one teacher. Standardized tests, then, as Edward Haertel (1986) states:

> tend to be unsuitable for measuring educational achievement as distinct from student competence, because they sample broad subject domains and are unlikely to match closely the curriculum in particular classrooms at particular times. Their breadth of focus makes such tests more sensitive to student individual differences beyond the teacher's control and less sensitive to the quality of current instruction.

For an accurate, fair view of student achievement and teacher performance, test score influences other than teaching quality have to be accounted for and controlled. Although this is no easy task, methods continue to be devised to reduce the competence factor and provide a less obstructed view of achievement. Two kinds of methods have been proposed to measure teacher effectiveness: one is a simulation method, which sets up classroom teaching situations with controlled content and time for teaching; the other is a naturalistic method, which uses actual classroom test scores (carefully controlling for nonachievement factors) as well as other classroom materials and evidence of teacher performance.

A well-known simulation method is the Popham-McNeil-Millman (PMM) approach. In a classroom situation, the teacher is provided with an instructional objective (specified in measurable learner behaviors) and a sample test item to show the teacher how the objective is to be assessed. The teacher is allowed a presentation time of fifteen minutes or more, with background information supplied if he or she is unfamiliar with the material. The teacher is given planning time--usually an hour or two--to work up an instructional plan. Then, as W. James Popham (1971) has described the method,

> a small group of learners (6-8 students), randomly selected from a pool of appropriate learners, is instructed by the teacher. After the instruction a posttest, not previously seen by the teacher but readily inferrable from the instructional objective and sample test item, is administered to the students. The pupils are also asked to supply an affective rating of the instruction, such as the degree to which they found the topic interesting. The performace of the students on the posttest and their affective ratings of the instruction serve as an indication of the degree to which the teacher is skilled at this particular task,

namely, the accomplishment of pre-specified instructional objectives with positive learner affect.

The advantage to this method--and the major reason for its being developed--is to create a fair way of comparing teachers' performances. When teachers are pursuing different instructional goals, it is impossible to make meaningful comparisons and ratings. If five teachers teach the same objective, it is more likely that a ranking based on performance will result (Popham 1971).

Inconclusive reliability. To claim to be reliable as a measure of student achievement, such a simulation method has to return a relatively stable judgment on achievement and the teaching inferred from the achievement. That is, it must show reliable effects across topics and different groups of students. Unfortunately, the reliability of this method is inconclusive (Glass 1974). The test does control for background knowledge of a topic, putting each teacher at the equal disadvantage of having to present a new topic. But is it reasonable to assume that good teaching does not involve background knowledge of a subject? Such an assumption might lead to a misleading division between a teacher's knowledge of the field he or she teaches and "background" knowledge of teaching techniques. (There may also be an implicit assumption that students respond largely to presentational technique rather than to the interest the teacher generates in subject matter.)

Moreover, to get sufficient evidence for rating teachers, the simulations would have to be performed not once or twice, but repeatedly. One calculation is that the method would have to be repeated "across ten different instructional topics with ten different pupil groups before the average score for a single teacher attained a reliability above .80" (Glass 1974). Repeated across the various disciplines in a secondary school, this method would be extremely costly.

Although simulation methods are helpful in staff development, new approaches may be necessary to make them useful and fair for rating teachers.

Naturalistic approaches. Noting the problems in using standardized tests and simulation methods, some researchers have proposed multiple-measure approaches using classroom data to advance product-oriented teacher evaluations. Two such models show that this naturalistic approach is usually a hybrid approach, as well, using several sources of evidence for student achievement and teacher effectiveness.

Glass (1974) has proposed a loosely structured evaluation system that uses trained classroom observers, student evaluations, and collateral data. His suggested "Observational-Judgmental

System" emphasizes the specificity needed in observations. Pupil evaluations of teachers would be confined to judgments about the learning climate in the classroom. Special attention is given to instances in which the observers' ratings of teacher rapport with students do not correspond to the students' ratings. Pupils' views, then, can serve as a partial correction to the rating process. Collateral data refer to minimum competency testing for teachers to eliminate the "math teacher" who can't graph a linear equation.

A more detailed proposal for evaluations, made recently by Edward Haertel (1986), expands the use of collateral data to include teaching artifacts (such as inclass tests or handouts). Haertel's model also includes specially controlled testing to try to link achievement and teacher performance. Regardless of students' test performance, portfolios of student-achievement evidence would also be examined in the Haertel approach. These materials might include completed practice tests, regular classroom tests, samples of student written work, homework papers, or teachers' observations of students. Additional information might include student attendance records and records of special remediation.

Establishing an appropriate, controlled testing procedure is more problematical, Haertel points out. It is important to take pains to make sure that teachers address the same learning objectives, teach comparable students, and have access to comparable school resources. The test items themselves would be developed using Item Response Theory. Combined in pretests and posttests, such items would allow tests that are focused, reliable for different levels of difficulty, and scored on a common scale.

Haertel states that two years of pilot studies and trial implementation would precede the first year of pre- and post-tests for evaluation purposes. Pilot studies in the first year would develop norms for student development. Standards can be set from pilot data and input from teachers, administrators, and students. After initial standards are established, they would be monitored for another one-year trial period and revised if necessary. The makeup of student groups tested must be controlled, as well. Three groups of students would be excluded from the scoring: students who were absent frequently, those whose posttest scores were markedly different from efforts on practice tests, and those who performed poorly despite the teacher's special efforts to help them. Under these conditions, a teacher would fail an evaluation if the class's posttest scores were below standard.

One advantage to Haertel's proposal is its combination of criterion-referenced and norm-referenced testing, made possible by creating test items from the curriculum of the school and then standardizing performance expectations. Ideally, criterion-referenced tests measure student achievement more accurately than

do norm-referenced tests, such as standardized achievement tests (which, as we have seen, make no distinction between competence and achievement).

Clinical Supervision Models

Clinical supervision is both a philosophy of teaching and a series of evaluation steps. As a philosophy, it emphasizes the collaborative nature of supervision; indeed, it is more a supervisory than an evaluative method. The focus is on teacher motivation and improvement rather than on summative judgments of quality. Morris Cogan, who formulated the clinical approach, held that teachers have a professional body of knowledge that can be adequately evaluated only by equally knowledgeable supervisors. The supervising of teachers, Cogan noticed, was akin to the supervisory practices of doctors--thus the name "clinical supervision." Noreen Garman notes that the relationship between supervisor and teacher is at the heart of Cogan's philosophy:

> The "clinic of the classroom" was a way of describing the activities where supervisor and teacher work together *every* day for a prolonged period of time (a practice not generally done in other forms of supervision). Cogan also liked the term "clinical" because it had an element of realism associated with it, as well as referring to someone who is trained to observe and analyze events in an empirical fashion (Garman 1986).

Collegiality, then, is essential to the assumptions behind clinical supervision. Thomas J. Sergiovanni (1982) notes five other assumptions of clinical supervision models: First, teaching is a complex set of activities that requires careful analysis. Most forms of evaluation tend to oversimplify the nature of teaching by starting with predetermined criteria. Clinical supervision, by contrast, derives issues from the teaching situation and takes seriously the teacher's analysis. Second, supervision is a "partnership in inquiry," with the supervisor as a more experienced practitioner instead of an aloof expert.

Third, the purpose of clinical supervision is to assist teachers to modify existing patterns of teaching in ways in which the teacher desires. Thus, supervision responds to teachers' instructional needs rather than administrative needs. Fourth, the supervisor's job is to help the teacher choose goals, to shed light on teaching issues, and to aid in the teacher's progress toward goals. And finally, effective clinical supervision increases the teacher's desire for and skills of self-improvement.

The stages of clinical supervision begin, as do goal-setting models, with preobservation conferences. Then comes the

observation itself, followed by the supervisor's analysis of the data gained from the observation and a strategy to improve the teacher's performance. A feedback conference involves the teacher and supervisor analyzing and interpreting the data. The teacher then decides on alternative approaches for the future with the concurrence of the supervisor (Cogan 1973, Acheson and Gall 1987).

The problems in implementing clinical supervision, however, bring its philosophy into conflict with real-world teacher evaluation. Clinical supervision cannot work if administrators perform traditionally as evaluators. Instead, it requires the supervisors to be colleagues rather than part-time evaluators. The time constraints and the lack of knowledge principals often labor under may turn their attempts at clinical supervision into mere mechanical steps. They may try to have the appearance of clinical supervision without the substance. Garman observes that this orientation has produced supervisors who simply go through the motions: "Itinerant supervisors often report, 'We are *doing* clinical supervision in our school' (meaning we are following the plan of the method) or more direct 'I am using the cycle on a group of teachers'" (Garman 1986).

Some schools, however, are finding that clinical supervision can be effectively implemented using peer supervisors who share responsibilities and may observe each other. Such an approach has been proposed and used under various names: peer supervision, peer coaching, collegial evaluation, collegial supervision. This variation of clinical supervision has two salient traits. First, information obtained by collegial supervision is purely formative and is shared with the principal only if the teacher who was observed chooses to do so. Second, participation in the process is voluntary, and teachers may choose their own partners for coworkers (Ruck 1986).

In a situation where the principal and teachers share responsibility for supervision, the principal could conduct summative evaluations. Separating formative and summative supervision in this way has been reported to improve school climate and teacher performance if each level is willing to cooperate and coordinate goals.

Conclusion

With the practical uses of evaluations divided between formative and summative, researchers and practitioners alike are now seeking ways to provide accountability without resorting to repressive control over teachers' professional lives. The paradox remains that teacher improvement is linked to professional development rather than to accountability. However, schools need both formative and summative strategies to serve the needs of teachers, students, and the public. The problem, then, becomes

how to provide both formative and summative procedures--both evaluation and supervision--in the practical setting of already overburdened schools.

Fortunately, some systems have successfully synthesized these two purposes. According to the Rand Corporation study on evaluation practices, the most successful systems pay attention to four critical factors in running their evaluation programs. First, they are committed--in resources as well as word--to their evaluation process. Second, they ensure that evaluators are trained and competent. Third, they emphasize collaboration between teachers and evaluators in the process. Finally, they use an evaluation process that integrates general goals with a teacher's specific instructional strategies (Wise and others 1985).

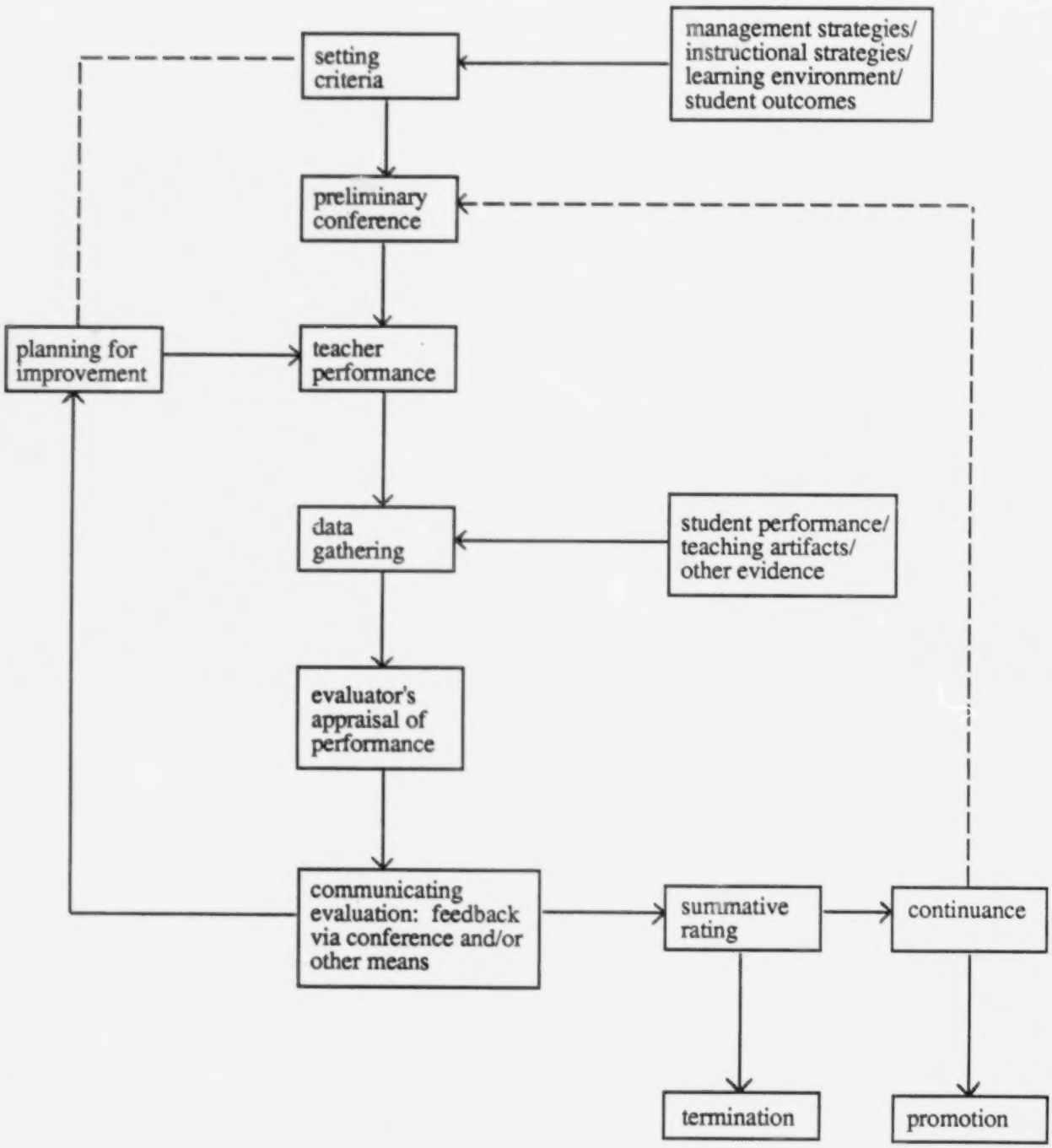
Based on the research into successful teacher evaluation practices, it is possible to ouline a teacher-evaluation model that combines the best and most needed features of other models. This composite model includes the evaluation contexts, teachers' interests in formative systems, evaluator's concerns, and the use of data relevant to improving and rating teaching.

Some models focus on one or another aspect of this process. Most concentrate on the stages surrounding classroom observations. As much research indicates, howeve,r school districcts can pay too much attention to observation without adequately preparing for it or following it up. There is also the summative dimension of evaluation to be considered: It poses a hard fact of life that research may ignore but that school districts must contend with.

When seen graphically, as in the accompanying diagram, teacher evaluation is the focal point of considerable energy. Implicit in setting criteria is a philosophy of teaching. Stretched over multiple evaluators (and teachers), the process could be adapted to many philosophies and practices of teaching. It gathers in student performance, teacher performance, and administrator performance. The process raises questions of valid data and of reliable and consistent interpretations of teaching effectiveness. It give rise to issues of promotion and competence, and also demands attention to claims to professionalism among teachers.

Above all, it is a process that must somehow balance accountability and development. In this conception, the development process both precedes and follows the summative ratings. Research and practice indicate that the formative and summative tasks are not only necessary but can also be complementary: development leading to a "test" of competency; the "test" leading into further development. Thus, as in the diagram, the cycle of evaluation can be a continuous professional development.

Evaluating Teachers for Accountability and Development: A Composite Model



The following chapters investigate in detail the roles of evaluators and teachers, their cooperation, and the variety and use of data in meeting both general standards and teachers' individual needs.

Chapter 2

The Evaluators: Who They Are and What They Do

Most evaluators are not specialists in evaluation. They are administrators who are compelled by district requirements to observe and rate teachers. One estimate is that 80 percent of instructional supervision is carried out by line administrators (McGreal 1983).

Although administrators can make good observers if they are aware of teachers' problems and have teaching experience themselves, the evaluation system they must work with and their own positions often prevent them from gaining teachers' confidence. Generally, they must use summative criteria in evaluation, designed to detect incompetence rather than to provide feedback for improvement. Their bureaucratic behavior, then, exacerbates their bureaucratic roles as teachers' bosses.

Administrators themselves feel discomfort at this situation. Being thrust into the dual roles of staff developer and evaluator brings down on them the whole weight of the dilemma surrounding evaluation: How can evaluations both improve and rate teachers? Where schools monitor teachers' performances continuously, they can enforce the minimum standards of performance. This is clearly an appropriate role for an administrator. If a teacher appears to overreact to student misbehavior, for instance, the principal could catch the problem soon rather than wait for a formal evaluation time. Effective instructional leadership is, in part, just such continuous contact with teachers--a kind of "management by walking around."

Separation of Summative and Formative Tasks

Teacher improvement instruments, however, require that a special set of procedures be established. Some districts are setting up such procedures for supervision, keeping them separate from the evaluative (that is, summative) tasks most often done by administrators. In such a separation of duties, the evaluation system benefits from having the most knowledgeable advice on teaching. Evaluators' competence, after all, is probably the most difficult part of an evaluation process. Administrators may also benefit by being relieved of having both to evaluate and advise teachers. They are less pressed for time by delegating teacher supervision and advisement. Also, to a lesser extent, they do not experience the possible role conflicts accompanying the dual tasks.

Each of the four districts studied indepth by the Rand

Corporation researchers used a differentiated staffing structure that separates formative and summative evaluations (Wise and others 1985). In each case, committees of teachers and administrators chose the teacher-experts on the basis of their skill in teaching and their interpersonal skills. All four districts also provided inservice training for evaluators, covering evaluation goals, procedures, and techniques. One district gives principals a two-week workshop every summer that includes Madeline Hunter's Instructional Theory into Practice techniques, clinical supervision skills, and rating methods. During the school year, these same principals attend monthly seminars reviewing and expanding on that material.

In addition, all four districts in the study have mechanisms for checking on the accuracy of evaluators' reports about teachers. The evaluators must defend their ratings in specific detail. Even when evaluators' reports fail to catch unsatisfactory teaching, the districts also have review-of-services or school-performance assessments that are meant to ensure minimum standards are met.

The division of labors differs somewhat from district to district in the Rand study, but all four districts report considerable success in maintaining a minimum standard of teaching and extending gains in teacher competence. In two districts, principals evaluate teachers and initiate probation and remediation procedures when necessary. Once probation begins, however, expert teachers provide the help to those needing improvement.

Another district operates a peer-adviser program for first-year teachers. Experienced teachers receive stipends and released time for their services. In yet another district, both principals and teacher leaders evaluate and offer advice, but a large pool of senior teachers coach teachers and set evaluation criteria.

A districtwide pool of supervisors, all with extensive and successful teaching experience, has been used as part of the School Improvement Program (SIP) in Pittsburgh. Working with the principal, the supervisors determine the instructional needs of each school through student achievement data and other sources. Then, they focus attention and time on specific areas in a school--a particular department, for instance, or the content-area skills of a group of elementary teachers. They can also focus on individual teachers, grade levels, individual students, instructional techniques for an entire school faculty, or whole programs. Developing yearly long-range goals for each school, the SIP supervisors can concentrate on short-term action plans, written every two weeks. They work closely with the principal of each school in coordinating and keeping current the goals for instructional improvement (Bickel and Artz 1984).

As a way to maintain teacher confidence in evaluations, then, and to provide administrators additional time for other duties, some differentiated staffing arrangement seems advisable. Once the roles are divided, though, experience has shown that two questions remain to be considered: what tasks the supervisor or evaluator must do, and how best to approach the teacher-supervisor relationship.

Expertise Needed by Evaluators

Decisions made by evaluators or observers as they approach their tasks largely determine the value of the observations and analyses for teachers. Although this statement may be obvious to most people, only the evaluators and observers themselves usually appreciate how difficult it is to decide what to look for and how to rate teachers. Not all observations may be valid nor, as we have seen, reliable. Different observers may look for different indicators of a teacher's competence. Some may focus on interactions with pupils, others on the teacher's classroom management, still others on the amount of preparation or the teacher's ability to stick to lesson objectives.

What the evaluator notices as significant data will depend somewhat on the model of evaluation used. The clinical supervision models, for instance, compel evaluators and supervisors to form a plan of observation with the teacher, concentrating on the teacher's perception of areas needing improvement. However, even evaluators who are operating with a narrowed focus may notice behaviors or may raise questions that the established criteria do not address. How does the evaluator know whether these observations are significant to teaching and learning?

One of the problems with generalist observers such as principals is that they often make unsystematic observations or base their judgments on vague, poorly defined criteria. Often, their criteria are drawn from the vague categories used in common-law models. For example, what evidence does the observer have of a teacher "developing good working relationships among students" or of a teacher helping "to carry out school policies and regulations"? Clearly, to be of use for teacher improvement or ratings, these lines on an evaluation form will have to be made more specific. Even before specifying them, though, those involved in the evaluation process will need to decide whether these traits even apply in a particular case.

Definition of Terms

Because research and evaluation instruments often contain technical terms for the qualities of teachers, it may be helpful

to clarify some commonly used terms. What is meant by *teacher competency*, for instance, and how might it differ from *teacher effectiveness*? Although these usages may differ from source to source, the research team of Medley, Coker, and Soar (1984) have defined four terms, each of which relate to what evaluators may be looking for:

> *Teacher competency*: a specific knowledge, ability, or value...that a teacher either possesses or does not possess, which is believed to be important to success as a teacher.
>
> *Teacher competence*: the repertoire of competencies a teacher possesses. The more competencies a teacher possesses the more competent the teacher is said to be.
>
> *Teacher performance*: what the teacher does on the job; it is defined in terms of teacher behavior under a specified set of conditions. How well a teacher performs depends in part on how competent the teacher is...and in part on the situation in which the teacher performs.
>
> *Teacher effectiveness*: the results a teacher gets. It is defined in terms of what pupils do, not what the teacher does or can do.

Besides these four, there is a fifth kind of evidence that may be considered in teacher evaluations--the teacher's personal characteristics, such as mannerisms and manners of speech.

Teachers may be evaluated on any of these characteristics. But evaluators should be both diplomatic and clear in deciding what data are fair game in evaluations. Evaluations of minimum competence will concentrate on individual competencies and their strength as a group in the teacher's performance. These are the categories considered when a teacher is first certified, answering the question "Is this person qualified enough to teach?"

The last category--teacher effectiveness--is, strictly speaking, a measurement of what the students do as a result of being taught. Presumably, we can trace student outcomes to a teacher's work, but as we have noted in regard to process-product models, tracing back from effects to causes is often problematic. Thus, it is *teacher performance* that is being evaluated for most teachers on the job rather than *teacher outcomes*.

The Dimensions of Effective Teaching

Once an evaluator sets out to examine a teacher's performance, two essential questions of measurement arise. The

first question involves developing a measurement instrument for the evaluator's (or school's) purposes: What distinguishes effective teaching from ineffective, if we are observing classroom teaching? In answering this question, an evaluator must give an indication of the dimensions of the performance to be evaluated. Those dimensions are usually set down in common-law models, but they are vaguely stated.

The second question concerns the definition of the task: What is the task the teacher has set, and is the teacher performing it well? The teacher and the evaluator must agree on the job to be done while the evaluator observes.

Defining the dimensions of an evaluation is probably not a process that an evaluator can do alone. However confident the evaluator may feel that he or she knows which areas of teaching are most important, he will probably need help at some time. If the present evaluation system needs reforming, an evaluator will want to consult various sources of information.

Theories of teaching. One source of information is a plausible theory of teaching. A theory is useful for interpreting a teaching performance and drawing conclusions. Theories abound. To be of use, however, a theory must be simple.

Medley, Coker, and Soar propose a theory based on three levels of teaching, all occurring simultaneously in a classroom: environmental maintenance, implementation of instruction, and individualization. The one indispensable task of a teacher, in their view, is "to create and maintain a classroom environment favorable to learning." This is a valuable basis on which to assess teachers' behaviors because if the learning climate is favorable, pupils will learn something regardless of other factors. A second basis for assessment is the teacher's implementation of the lesson plan. Has the teacher made the objective clear to the students? The third basis for evaluation is the attempt the teacher makes to adapt the lesson plan to keep students involved.

Other models may suggest other strategies, based on what they hold to be necessary for effective teaching. *Models of Teaching* by Bruce Joyce and Marsha Weil (1986) provides some specific models of teaching that suggest what to look for in observing teacher performances. Hunter and Russell's (1977) model of lesson design is also very suitable for evaluators.

Consensus of practitioners. Evaluators can also use local consensus about effective teaching as a guideline. A group of practitioners who will speak from their experience can be drawn from the school or district itself--ideally from the group of teachers who will be evaluated. Medley and his colleagues used the consensus of a group of teachers to develop lists of

competencies and of accompanying behaviors that characterize successful teaching. The example in table 2 translates two general competencies into specific, observable behaviors.

Table 2
Examples of Effective Teacher Competencies and Behaviors

Competency Area	*Teacher Behaviors*
1. Organizes pupils, resources, and materials for effective instruction	a. Selects goals and objectives appropriate to pupil need
	b. Matches pupil with appropriate material
	c. Gathers multilevel materials
	d. Involves student in organizing and planning
2. Demonstrates ability to communicate effectively with pupils	a. Gives clear directions, understood by pupils
	b. Pauses, elicits, and responds to pupil questions before proceeding
	c. Uses a variety of methods, verbal and nonverbal, to deliver instructions

(Source: Medley and others 1984)

So many lists have been compiled over the years that administrators would save time and effort by adapting an already developed list--perhaps by submitting it to a group of teachers for revision.

Research findings. Less comprehensive than theory and less immediate than consensus, a third way of specifying dimensions of good teaching is through research on effective teacher performance. Similar to consensus studies, research provides some tested ideas on teaching. It also provides some tentative, untested methods that a district might benefit from trying out. Summaries of teacher-effectiveness research often provide suggestions for application. Even though research can provide guidelines for an evaluation of teaching, it can never provide one best way to teach.

Measuring Performance

Having used these various sources to determine what dimensions of teaching should be evaluated, evaluators next need to find how a teacher is performing in each area. Accurate measurement of performance requires that there be some limitations on what the evaluator observes. Even though someone can collect an accurate, objective record of what occurs in a classroom, it may still not be a valid measurement of a teacher's performance. The record may leave out the factor of what the teacher was trying to accomplish. "It is necessary," believe Medley and others,

> either to set a task for the teacher to perform or to let the teacher decide what she is trying to do and arrive at a clear understanding of it....Only if we know what the teacher's purpose is can we assign positive weights to relevant behaviors that reflect best practice for accomplishing the teacher's purpose, zero weights to ones that are irrelevant, and negative weights to relevant behaviors that do not correspond to best practice.

Thus, it is important that evaluators and teachers confer on teacher objectives.

One common objection is that evaluator/teacher conferencing before an observation makes the teacher less "natural," more likely to perform for the evaluator rather than act as he or she might when not observed. This objection is probably unrealistic, however. Most teachers, even when surprised with an unexpected visit, will try to do their best anyway. Moreover, when the evaluator has no idea of the teacher's objectives or classroom management problems, the evaluation may be more punitive than helpful.

A clear task definition and a clear set of assumptions about what constitutes good teaching are both crucial. Even where teacher and evaluator have implicit assumptions--that is, when they *assume* they share the same ideas--the dissonance between intentions and perceptions may lead the evaluator to the wrong impression and the teacher to a negative view of evaluation.

The validity of evaluations appears to depend, therefore, on two essential exchanges of information: (1) from evaluator to teacher, on the general bases of effective teaching; and (2) from teacher to evaluators, on the objectives intended during any given observation.

Relationship between Evaluator and Teacher

The issue of an evaluator's expertise encompasses more than

the performance of the evaluator (or the evaluation system) in judging a teacher's competency or effectiveness. Evaluation also deals with human relationships. Because the experience of being evaluated is often emotional and occasionally provocative, teachers want to believe in the worth of the evaluation and the evaluator. Evaluators should have something valuable to offer and be willing to understand the teacher's obstacles, as well. Evaluators have both a social niche in the school culture and a personal influence. Both the organizational influence and personal influence affect an evaluator's ethos and credibility.

An Unsupportive School Culture

Social psychologists have pointed out that the school environment sets out norms for the behaviors of teachers and administrators. The workplace culture in a school affects how people act: how they teach, learn, and evaluate performance. Because these are norms of behavior--that is, they deal with how the school actually functions rather than how someone thinks it should function--they are not so tidy as we might hope. Schools, for instance, as hierarchical organizations, are susceptible to outside pressure. To maintain their coherence, however, they usually adopt a relatively "loose" structure: The parts of the organization function autonomously (with teachers having unique control over their classrooms). Policies conceived at the top are easily derailed as they move into the faculty.

Among the faculty, there is an absence of consensus and considerable pride about individual pursuits, particularly on issues of values and objectives. Lacking consensus on decision-making matters, the adults in a school must resort to bargaining and compromise to implement plans. Although principals may exert influence, they must rely on the consent of the staff and must be willing to share power with teachers, who often guard their own professionalism.

In this sort of environment, teachers are mostly on their own when they want to improve their instruction. Pressure or encouragement from other teachers to improve their teaching is rare. Unless an effective cadre of peer supervisors is present, teachers as a group tend to tolerate or approve a wide range of teaching behavior as part of their professional ethic. Arthur Blumberg (1983) has noted that, as a rule, teachers can be as industrious or effective as they wish, with little group pressure to be better than they are. Of course, this is not teachers' conscious choice but a result of the school's social organization. To be criticized by other teachers for their performance, Blumberg notes, a teacher has to be so ineffective as to make life harder for other teachers. For instance, a teacher whose classroom is too noisy will receive complaints from other teachers and perhaps some assistance in classroom control.

The evaluator or supervisor is separated from teachers' roles and rights, yet must contend with the same lack of support for teacher improvement that teachers experience. For one thing, supervisors who attempt to change teachers' practices must do so at the option of the teacher. Observing a classroom teacher does not necessarily lead to having influence on the teacher. Thus, teachers often put limits on the extent that supervisors may intervene in "teacher territory."

Supervisors are also constrained by the organization of the school when the school provides no rewards for better teaching. Unable to reward teachers, supervisors have to depend on more intrinsic motivations. As Dan Lortie (1975) has pointed out, the rewards of teaching are primarily intrinsic anyway. But improvement strategies are often hard to get started without a teacher's early sense of being able to gain something from the strategy. Consequently, the work of the supervisor may be "a slow process, through which changes, when they occur, may be barely perceptible" (Blumberg 1983).

Attitude of Reciprocity

It seems reasonable, then, that a supervisor should work with teachers individually, cultivating knowledge of the teacher's goals and communicating a sense of the worth of the teacher's work. The evaluator must, at the least, appear to maintain the "logic of confidence" in the teacher's role in the school. This confidence--assuring the teacher that his or her autonomy and experience will be respected by school authorities--is characterized by the lack of direct interference in a teacher's work and a sense of reciprocity in an evaluator's attitude. The idea of teachers' professionalism, too, is an expression of the maintenance of mutual respect (Meyer and Rowan 1978).

The requirement of reciprocity--the feeling of mutual respect between teacher and evaluator--has been said to be at the heart of the evaluation process (Blumberg 1974). Observers' attitudes often come through more strongly than they may realize. Tom Bird and Judith Warren Little (1985) have drawn up a contract-type list of attitudes and acts that can create a climate of reciprocity in the evaluation setting:

The Requirement of Reciprocity

- The observers must *assert* the knowledge and skill needed to help a practitioner of a complex craft. The least assertion which can be made in observation is something like, "I can make and report to you a description of your lesson which will shed new light on your practices

and thus help you to improve them." That is the least assertion that can be made. It is a substantial assertion of knowledge, skill, and discipline. The question is what training and experience, either in teaching or in observing, would permit the observer to make the assertion in good faith.

- The teacher must *defer* in some way to the observer's assertion, for example, by allowing the observation, by teaching under scrutiny, and by listening carefully and actively to the observer's descriptions, interpretations, and proposals. The question here is, What prior knowledge or experience does the teacher need to grant the observer's claims to knowledge and skill, and thus to participate in the observation in good faith? How could the observer have attained, in the teacher's eyes, the stature which must be asserted in the observation?

- The observer must *display* the knowledge and skills which s/he necessarily asserts. The observer must make a record of the lesson which is convincing and revealing to the teacher of the lesson, or propose an interpretation of the lesson which can make sense to the teacher, or must offer feasible and credible alternatives to the practices which the teacher used. How can the observer gain and refine those skills in practice?

- The teacher must *respond* to the observer's assertions, at least by *trying* some change in behavior, materials, role with students, or perspective on teaching. Such changes are known to require effort, discipline, and courage, but if they do not occur then the observation was fruitless. Here, the requirements of observation become practically circular. The requirement of reciprocity in observation is not met without change on the teacher's part; changes in teaching behavior, materials, roles, and perspective are difficult to make without close support such as observation and feedback. The observer and teacher must start with modest efforts at which they can succeed, meet the requirements of their relationship, and then build on those gains.

- The observer's performance must improve along with the teacher's and by much the same means: training, practice, and observant commentary from someone who was present. Observation cannot be simpler than the teaching it supports. If the observer does not advance with the teacher, the observer's assertions of knowledge and skill gradually are falsified. And the central premise of observation--that mutual examination of professional practices is necessary and good--is shown

to be a lie. (Bird and Little 1985)

As this statement makes clear, evaluators work most effectively when they have knowledge of teaching--both as a profession and as each teacher may practice it. Personal qualities of the evaluator certainly enter into the picture--their trust level, their patience, and their persuasiveness. But the impressions evaluators make result largely from their professional traits: their credibility, developed through their own experiences with teaching; their knowledge of each teacher's goals and unique difficulties; their track record as a supervisor and advice-giver; and their ability to model new ideas or techniques for teachers (Duke and Stiggins 1985).

Training for Both Evaluators and Teachers

In short, evaluation requires as much clarity about objectives and methods as teaching itself does, and fully as much interpersonal skill. The reciprocity of responsibilities means that a functioning evaluation system should probably provide essentially the same training for teachers and supervisors. Thomas McGreal (1983), drawing on his observations of effective and sham evaluation systems, believes that all participants in a system must have the same training:

> With the exception of additional time spent with supervisors on their responsibilities in the goal-setting conference, on observation techniques, and on conferencing and feedback skills, administrators and teachers should initially receive approximately the same training.

He offers a general outline of a training program for an entire staff and for supervisors. Flexible in format, the program is suitable for inservice, perhaps best conducted by someone from outside the district. Specialists can be brought in to cover followup sessions; subsequent inservices can address other aspects of the teaching-learning process or teacher-supervisor relationship. Indeed, a whole crop of inservice topics can be generated from this seed. (For an outline of this training program, see the Appendix.)

Chapter 3

The Teachers: Concerns and Participation

In the mid-1970s, Arthur Blumberg wrote of a "cold war" between teachers and evaluators. In many districts, the same tension and doubts about the value of evaluations persist.

One teacher, quoted by Duke and Stiggins (1986), complains that the principal showed up for the evaluation twenty minutes late and stayed only half an hour: "Did the principal know I ran into trouble and had to change plans midstream? Why did the kids choose that time to behave as they did? Did the principal realize that every day is not like this?" At the postobservation conference, the principal's comments were (as ever) flattering. The teacher was relieved but also mystified: "It's always the same--I never understand why I get nervous!"

A second teacher, who had received complaints from parents about her teaching, had a notably awful day when the principal finally came to observe:

> On my observation day, my principal came in early, just as two kids started fighting; three others were throwing paper. That was just the beginning. Nothing seemed to go well from that point on. She stayed for ten minutes and left with a scowl on her face. At the end of the next day, during our postobservation meeting, she said these were the problems she saw in my class: students were undisciplined, I was poorly organized....The list continued and I nodded as she reviewed each problem. Now she wants me to write out a plan for making changes, but I have no idea where to begin. What I need are some concrete ideas, but no one is available to help, particularly the principal. She thinks all you need to do is tell teachers what's going wrong and have them write out a plan. What I need is real assistance, not just a bunch of complaints.

In the first case, a strong teacher's evaluation becomes an excuse for hollow-sounding praise. *Why* was this teacher particularly good? What, in particular, did the principal see that was excellent? Was there any room for improvement? Did the principal notice the change of plans? Did the principal's tardiness affect the evaluation?

In the second case, a marginal teacher is given an evaluation that points out the obvious. As in the first example, the evaluation raised more questions than it answered. Where should this teacher begin in getting control of the class? Are there techniques useful for rowdy pupils? What is the trouble with her

ganization--specifically?

Both evaluations were structured in formats with econferences, observations, and postconferences. Both may be edible and reliable (any number of observers who saw the same sons may have reached the same conclusions). However, both aluations flunked the test of *usefulness.*

Teachers' views are not unrelieved hostility toward valuation. Most teachers, with some important reservations, upport evaluations if they are useful in improving teaching. ndeed, teachers' views often coincide with administrators' views f the barriers to effective evaluation systems.

As part of their study of evaluation practices in four 'acific Northwest school districts, researchers Stiggins and Bridgeford (1985) assembled teams of educators--each team having a listrict administrator, principal, and teacher--to consider the ssues surrounding formative evaluations. The conference participants produced a list of common barriers to evaluation for teaching improvement. Foremost was the evaluators' lack of training in rating teacher performance knowledgeably and in communicating with teachers about the results of observations. In other words, credibility of the evaluators was the key problem. In both examples presented above, the principals could easily be construed as shirking the duty of offering advice. Instead, they provided general judgments.

The other barriers noted by conference participants may be easily recognized as common to many districts' evaluation systems:

> There is insufficient time for both evaluation and follow-up....The competing demands of education frequently push evaluation to a low priority status.
>
> The process(es) for linking staff development and teacher evaluation is (are) not clear. [Districts] lack a clear goal for formative teacher evaluation (i.e., an image of the desired system) and a plan for achieving that goal....Despite an important emphasis on protecting the due process rights of teachers, evaluation systems lack a similar commitment to promoting professional development.
>
> Unclear or unacceptable performance criteria, combined with lack of teacher involvement in developing performance criteria and infrequent and superficial observations, tend to breed skepticism among teachers about the value of results. The adversarial relationship between districts and collective bargaining units also breeds distrust. (Stiggins and Bridgeford 1985)

Teacher Participation and Program Success

The single most frequently mentioned barrier to effective teacher evaluation, it appears, is that teachers too often lack significant input and participation in evaluation systems. All too often, evaluation systems are bureaucratically rather than instructionally centered. Teachers may express their complaints about lack of participation in a number of ways.

One survey found that teachers viewed their evaluation systems as generally inaccurate, often because of overly subjective judgments on the part of evaluators. Furthermore, they felt that evaluations were unaffected by their efforts. The criteria used in evaluations were rarely shared with teachers, nor did teachers have access to the information collected in support of the evaluation (Natriello and Dornbush, in Stiggins and Bridgeford 1985).

Punitive and Unfair Evaluations

When all the evaluatory force is on judgment rather than problem-solving, teachers are likely to be defensive. They see evaluations that do not include their points of view as arbitrary and unfair. Arthur Blumberg (1974) surveyed experienced teachers about needs that are satisfied or unsatisfied by evaluations. The most negative evaluations, these teachers said, were those that viewed the teacher arbitrarily from outside their roles rather than from a teacher's perspective. Teachers said they felt evaluations were often punitive, inviting hostile interpersonal criticism from supervisors. One teacher commented that, when her supervisor found out she had done well on the National Teacher Exam, "she said that she didn't see why my classroom discipline wasn't better since I was so smart."

A second criticism was that evaluations were not fair: supervisors used inadequate information to judge teachers. For example, a teacher in the Blumberg study said that he was criticized by his supervisor for poor spelling when the words he had written on the board had been purposely misspelled. The supervisor had not bothered to ask.

These examples are more than simply instances of the judge falling asleep on the bench or of personal insensitivity from an evaluator. They indicate what many recent discussions of teacher evaluation have brought out--the absence of teacher input into the process of evaluating. Even more than personal insensitivities, teachers object to *professional* insensitivities in the evaluation process that make the evaluations inaccurate or wastes of time.

As the preceding chapter's discussion of the evaluator's role has already shown, the environment of teaching often provides no

ready support for improvement. Furthermore, teaching has few stages or plateaus to which teachers progress, as do other professions (say, medicine or law). Nor does teaching provide sure measurements of success, such as lawyers have in winning cases. Given a profession, then, that defies clear evidence of accomplishment, it is no wonder that teachers flinch at evaluations that do not appear to take their profession seriously.

"Remote Control" Governance

Unfortunately, much of the new enthusiasm for teacher evaluation at policy-making levels fails to pay attention to teacher input. Competency testing plans and merit pay proposals typically are based on standardized lists of what good teaching is, regardless of the context of individual teachers' goals, content areas, or student makeup. Several policies are counterproductive to improving classroom teaching, according to teachers surveyed by the Rand Corporation: (1) curriculum and testing policies that limit what can be taught and how, (2) policies that create paperwork and divert teachers' energies from instruction, and (3) policies that deprofessionalize teaching by excluding teachers' judgments about what constitutes appropriate teaching and learning (Darling-Hammond and Wise 1983). On the other hand, other research has shown that teachers accept the evaluation process much more readily when they have some significant influence over it--even when individual evaluations are negative (Natriello 1983).

What teachers are objecting to has been termed "remote control methods for governing education." According to Darling-Hammond and Wise, this aloofness from teacher input conditions supervisors to look for only a narrow range of behaviors in teachers. Teachers become frustrated when they realize that the standards they are willing to meet have become hair shirts that they must wear to meet minimum requirements:

> They feel they have no time for activities that are not geared toward discrete cognitive skills that will be tested on multiple-choice tests used for promotion purposes, tracking purposes, or accountability purposes. Teachers complain that they have been limited in the choice of materials they can use--that they are limited, for instance, to a single basal reader that doesn't meet the needs of all of their children. They cannot pursue topics of the children's interest because they are supposed to be on a particular page on a particular day or they are supposed to achieve certain objectives by the end of the classroom period. (Darling-Hammond and Wise 1983)

Thus, teachers often find themselves on the horns of a dilemma. On the one hand, the evaluation system may be bogus--an artificial process akin to playing a game: if it's a good day, you win; if it's a bad day, you lose. On the other hand, in districts that are heavily product-oriented or that attempt improvement by rigidly controlled standards, evaluations may be bureaucratic requirements rather than commitments to excellence in teaching.

Means for Involving Teachers

How best can the schools change an arrangement, then, that increases teachers' alienation, increases conflict, and offers little worthwhile assessment or flexibility?

Teachers surveyed on this question have recommended a number of valuable courses of action, each including teacher participation in devising and implementing evaluation plans. The teachers interviewed by Stiggins and Bridgeford, for instance, urged more collegial observation and self-evaluation through videotaping and goal-setting. Teachers repeatedly called for more frequent feedback and improvement-oriented criticism rather than vague generalities.

In a number of studies, teachers have emphasized the importance of schoolwide priorities for improvements in evaluation systems, rather than evaluators simply going through the motions. They have noted that evaluators need to use complete information that is specific and relevant to teachers' experiences. The consensus has been that "when the process of teacher evaluation is supportive and collegial, and when an organizational structure is more open than closed, allowing for teacher input and rational outcomes, the evaluation process will be perceived, by teachers, to be more positive" (Johnston and others 1985. See also Stiggins and Bridgeford 1985, Wise and others 1985, Blumberg 1974, Darling-Hammond 1986).

Reporting the results of another Rand Corporation study, McLaughlin (1984) makes two suggestions about involving teachers more responsibly in evaluations. First, school districts should designate expert teachers to observe and assist other teachers, particularly beginning teachers and those in need of special help. The experts should not only be excellent teachers themselves, she cautions; they should also be able to provide supervision and assistance *to adults*. Unlike children, adults must be motivated to learn by having new techniques or ideas connected to a practical need for them. Thus, expert teachers must be aware of individual teachers' needs and be flexible enough to provide alternatives. To ensure that they will have time for this attention, expert teachers should be given released time and/or additional contract time, the Rand researchers recommend.

Second, the school district can involve teacher organizations in designing and overseeing evaluation procedures. Traditionally, the role of management has been to enforce accountability; the typical union role, to afford protections. This distinction will be obscured if teachers begin to take more responsibility as a group for their professional standards--opening the door to collaborative control over teacher quality. Looking to the future, the Rand study sees teachers developing boards of professional standards, such as those that govern doctors or lawyers. Unlike the remote-control, bureaucratic approach, professional evaluation approaches will emphasize staff development and career incentives--issues on which school-improvement advocates and teacher unions may be able to find some common ground.

Other Human-Factor Suggestions

There are also other ways to make evaluations more "user friendly." One way--clarifying the performance criteria expected of teachers--could save teachers considerable confusion and spare supervisors frustration. Criteria pose problems when they are ambiguous, too general, or unrelated to teachers' actual practices. Often, they can focus on personal characteristics rather than instructional traits.

Which criteria are important enough to be generally used? And how should supervisors use them in relation to teachers? First, as an assurance of being *relevant*, the performance criteria should be reviewed by teachers--perhaps a districtwide council of master teachers--and endorsed by each teacher as relevant to his or her classroom. Criteria should be valid in each classroom environment, appropriate for content and instructional methods used, and flexible to allow the teachers a choice of strategies.

Next, to relate the general instructional program to each teacher's work, the criteria should *relate to student outcomes*, identified by teachers and principals together. Behaviors that make a difference for students are the important points: clarity of presentation, for instance, or direct instruction for some instructional goals.

The criteria should also be *practical for teachers*. Priorities must be set by evaluators and teachers to allow supervision to be accomplished in a reasonable time and with attainable goals. Finally, the criteria must be *clear, specific, and consistent* (within flexible limits) to ensure that the observation data will give teachers unequivocal feedback and a continuity of goals, regardless of who the evaluators may be.

Throughout the evaluation process, channels of communication

must be open between teacher and supervisor. The procedures are not exercises in fault-finding or in one-way communication. The evaluation process is a learning process for both parties. Settling on adequate criteria, researchers have noticed, is most often a reciprocal arrangement: one side monitoring for consistency, the other for flexibility and individuality of approach. Balances are attainable and, ultimately, the most useful approach.

Like communication, the teacher's freedom from unnecessary comparison is important to making appropriate criteria. Ranking teachers by proficiency, though it may occasionally be needed, as in master teacher appointments, most often simply subjects teachers to unwanted summative procedures. "After all," say researchers Stiggins and Bridgeford, "professional development, not criticism for its own sake, is the whole point of the system"--and the point of the careful development of evaluation criteria, we may add.

A teacher's responses to evaluator's comments, it has been found, are shaped in part by the evaluator's personal interactions with a teacher. Teachers tend to react negatively to more direct supervisory behavior, where a teacher perceives the supervisor as predominantly telling without reflecting or asking questions. Blumberg (1974) has found that teachers do not mind supervisors' telling, suggesting, or criticizing as long as they put equal weight on *asking* the teacher for information or opinions, or on *reflecting* on the teacher's performance. Passive supervisors ("He just sat there for twenty minutes and didn't give me any feedback later") are also perceived negatively.

Evaluators who talked more than listened (the *direct* style), Blumberg found, tended to approach evaluation as an issue of authority. In such a hierarchical approach, there is little place for collaborative problem-solving. Those supervisors who listened as well as gave advice were willing to let the problem determine the direction of events. They also tended to be aware of the teacher's need for formal recognition, as well as the intrinsic rewards that usually accompany teaching.

Conclusion

In sum, teachers' contribution to the procedures of evaluation, as well as to the outcomes, can be substantial. Their reception of evaluation as an improvement tool and as a rating instrument can make or break an evaluation system. The recurring theme of research studies has been that significant, real teacher participation in all phases of teacher evaluation changes an adversarial, irrelevant program into one of real use to teachers.

Using peer supervisors and master teachers may require

altering bureaucratic expectations for what evaluations will produce. The rate of evaluations, for instance, will change as observations become more frequent and perhaps of longer duration. The process will become more reciprocal, as well: evaluators being responsible for useful advice and sensitive interpersonal skills.

Making teachers fuller partners in evaluation can have gratifying results, as in one Minnesota district where teachers voted to continue funding evaluations as a high priority when the district's budget was trimmed, or in Washington State where teachers amended their collective bargaining agreement to emphasize more and even unannounced principal visits (McLaughlin 1984).

Chapter 4

Appropriate Data and Effective Feedback

So far, we have discussed the environment of evaluations, some common alternative models, and the interests of teachers and supervisors in the evaluation process. When people talk of evaluations, however, they usually are not thinking of these elements but instead of the classroom observations and perhaps of the evaluator's feedback to the teacher. Consequently, much has been written about observational techniques and the kinds of data collected.

The most common structure of evaluations has three stages, beginning with preobservation conferences, then moving to the observation itself, and finally having a postobservational conference.

Preobservation Conferences

Most supervisors consider the time on preobservation conferences well spent. Observations are more difficult and less helpful for a teacher when an observer enters a classroom unprepared. For many observers and teachers, a nondirective, informational conference is more effective than a goal-setting conference. In particular, supervisors want to know where a teacher is in a unit (beginning, middle, or end). They want to know what the teacher's objectives for the lesson are. Finally, they want to know what activities the teacher plans.

The preconference planning also gives teachers and supervisors time to review the data-collection procedures to be used. This is also the time for supervisors to ask teachers what else they should record--any specific problems the teachers want advice about.

Among the numerous suggestions that have been made for structuring the preobservation conference, some focus on information-gathering and others on goal-setting. Depending on teachers' individual needs, either purpose may be appropriate. Keith Acheson and Meredith Gall (1987), for instance, outline the following goal-setting process:

1. Identify the teacher's concerns about instruction.
2. Translate the teacher's concerns into observable behaviors.
3. Identify procedures for improving the teacher's instruction.
4. Assist the teacher in setting self-improvement goals.

5. Arrange a time for classroom observation.
6. Select an observation instrument and behaviors to be recorded.
7. Clarify the instructional context in which data will be recorded.

A goal-setting conference such as this requires that teacher and supervisor decide on strategies best suited to the outcomes a teacher wants, the techniques he or she plans to use, and a host of other situational factors. A contribution of the clinical supervision approach, this form of goal-setting may have the teacher collaborate with the supervisor in translating abstract concepts into observable behaviors. In the following dialogue between a teacher (T) and an observer (O)--drawn from a training manual prepared by the British Columbia Teachers' Federation (1986) for its Program for Quality Teaching--the teacher's concern is made into a specific focus for the observer:

T I don't think I explain things clearly.

O What's happening that makes you think so?

T Well, after I give an explanation, I usually ask questions about it. Sometimes they're just oral, but sometimes I give a worksheet or a quiz or something like that. A lot of the kids don't seem to get the point I've tried to make.

O Do you use any visual aids when you explain?

T Sometimes. But I'm not sure if they help....I've never checked it out. Maybe they do, but maybe what I'm saying just isn't clear enough.

O Do you think it might help for you to know exactly what you say in your explanation and what questions the students ask during and after your explanation?

T Yes...hey, it might help to know *which* questions too....Then I could compare that with the papers to see if a student who asked about a particular point handled that part of the work well. Yes, that might help me out.

O Fine. Then I'll collect verbatim data on teacher and student statements, noting which students ask which questions. After that, you might want to try a lesson using a diagram or an illustration, and we can see if that makes

a difference in student comprehension.

Observations and Data

Three dimensions of direct classroom observations recur in the research: the role of the teacher in observations, the challenge in focusing observations, and the selection of observation instruments. These dimensions form a view of observations as a structured and, thus, selective endeavor: structured by a prior framework, and selective through focus on detailed aspects of a teacher's behavior that experience and research indicate are significant in teaching. Like a literary critic, who reads a text carefully and selectively, the good observer is also a critic, but one who knows that he or she is watching a living text, one that generates its own ideas and ultimately must improve itself.

In the overall process of teacher supervision, classroom observation occupies only one phase; it is surrounded by preparation and followup and by the determination of objectives, standards, adequate instruments, and long-term developmental programs. It is one juncture in the web of teacher supervision, a highly important one but one that must be supplemented by other evidence of teaching performance and postobservational dialogue.

The teacher can participate in selecting or developing observational techniques. The feedback carries more weight with teachers if they have a hand in customizing the observational criteria to their areas of interest. Most teachers gain more from feedback related to a particular lesson's goals or activities. Moreover, when teachers help form the observation's methods, the data are more likely to be descriptive--more of a mirror held up to their teaching than value-laden judgments. Thus, for better reception of the results of an observation, there are compelling reasons for including teachers in preparing the instruments of observation. In the following example, a teacher and observer discuss how to observe the groups in a poetry class:

T I am not sure how it will work out, but I want to find out if homogeneous groups will produce a wider range of criteria than other methods have and if more students will participate in criteria selection.

O The answer to the first concern will be easy to obtain from the group reports.

T Yes, I thought that I would ask each group to have one of its members record and turn in the group's criteria. What can you do to help me check student participation?

O I could construct a verbal flow chart of each group to see which students were contributing and in what manner they were contributing.

T I don't think that I need that information from each group, but I would like it about the two slowest groups. Those groups will be composed of students who usually don't contribute.

O I can do that. Would actual verbatim data or an audiotape be better for you?

T No, the recorder would probably be too distracting for these kids. I just want to know who leads and who contributes in these groups. (British Columbia Teachers' Federation 1986)

Focusing the Observation

Active teacher participation in the planning conference can also help in focusing observations to record useful data. Focusing means choosing appropriate questions to interpret data. Observers need to use forms and recording instruments that allow them to describe accurately what goes on in the classrooms they observe. Even when observers have planned with teachers what they will see, they still need an instrument to map their observations, much as travelers in unknown territory need maps to orient themselves.

For many years, the common practice of observers was to observe without a plan, the theory being that an observer could be objective only with complete license to observe everything. Unfortunately, few observers *are* entirely objective. Without a narrowed focus on specific teaching activities, observers tend to see selectively, forming judgments that may have little to do with instructional matters. Quite often, unfocused observations say more about the observer's beliefs than about the teacher's behavior.

Goal-setting conferences help focus an observation, as can an agreement between teacher and observer about their philosophy of effective teaching. An observer who has a strong idea of what effective teaching looks like will often look for particular traits in a classroom performance: the teacher's use of engaged time, for instance, or the variety of instructional techniques used. Does the teacher allow opportunities in question-and-answer sessions for students to understand and apply what they are working on? Does the teacher provide a variety of ways "into" the

material--verbal, visual, kinesthetic? Are there puzzles, simulations, or stories? Does the teacher raise a question from a previous class, or do previous sessions seem to drop into a black hole, never to be referred to again?

The cardinal rule of observing is to focus on whatever behaviors and events might aid the teacher to teach more effectively. According to Ronald T. Hyman (1986), observers can be kept on task and aware of pertinent information by tying what they observe in the classroom to the nonobservational data also available to them. Nonobservational data include student achievement scores, attendance records, and written evidence of teacher relations with students. Observers should concentrate, too, on those activities central to teaching.

> Focus on what the teacher does and is directly responsible for, such as teacher questions, teacher reactions to student responses, teacher physical position in the classroom, and teacher selection of students to participate in the classroom interaction. Since these are the teacher's own actions, the teacher can change them directly (Hyman 1986).

Finally, observers should vary what they observe to cover a range of teaching skills. If a teacher has established a good classroom climate, for instance, the observer could look at the use of space in the classroom or at the nature of the teacher's questions, instead. Giving input about problem areas, after all, can be a strong motivator and can give professionals goals to work for.

Any developed criteria for teaching effectiveness can stimulate questions and structure the observation. Research into effective teaching has provided many such criteria.

Another way to focus observation is probably the one needing the most careful thought--that is, using a premade observational assessment guide. The advantage to using one of the packaged assessment instruments lies in their convenience:

> Being selective involves "taking a point of view," and the easiest way to take one is to choose an *observation instrument* from among the many our researchers have developed. An instrument has a built-in framework, a point of view or vantage point, as well as a set of rules for systematically observing and organizing data. In addition to guiding the observer in selecting what to observe, an observational instrument yields reliable and specific data which forms the basis for helpful feedback (Hyman 1975).

But the convenience of an instrument poses a problem, too. The

ready-made interpretation the instrument provides is someone else's interpretation, not the supervisor's nor the teacher's. The focus of the observation, then, must take priority over the instrument. The preobservation conference is simply a better guide to interpretation of data than a packaged instrument. Taken together, though, the personal information and the data supplied by the instrument can be highly persuasive and useful.

Types of Observation Instruments

Observation instruments come in different formats and produce different types of data. *Rating scales* are usually meant for ranking teachers and demand high-inference skills from the evaluator. For instance, on a criterion such as "the purposes of the lesson are clear," the evaluator may rate the teacher weak, below average, average, strong, superior, outstanding, or truly exceptional. Thus, there is an implicit comparison of one teacher to another in the rating scale--a fact the evaluator should consider in dealing with a teacher. For this reason, rating instruments are often used for summative evaluations and are not suited to formative evaluations.

Some rating instruments--those that have well-defined items--are more suitable than others for classroom observations. For instance, "states or writes down objective and plan of lesson for students" is a more well-defined version of the item in the preceding paragraph. It provides a teacher a clue about what behavior is expected and thus may be helpful in improvement-oriented evaluations.

The most persuasive data for teachers, though, are the most descriptive. Those systems that require complex inferences by the observer, such as rating systems, are less convincing because they are mediated by the observer's judgment. Writing descriptively, however, is a skill requiring training. Usually, the observer takes notes in some telegraphic style (short, heavily verbalized phrases) and expands them for the postobservation conference.

Category types of instruments sort classroom behaviors into classifications so that teachers and observers can focus on activities in one dimension of teaching. One category system is the Seating Chart Observation Records (SCORE), which record interactions on the basis of student seating charts. For instance, the Beginning Teacher Evaluation Study focuses on engaged time of students and success rates in interactions between first-year teachers and students.

Other systems, such as Acheson and Gall's System for Measuring Verbal Flow or Stallings's Teacher Interactions Form, focus on such fluid variables as off-task behavior and physical movements in group projects. According to McGreal (1983), these

variables would be hard for teachers to isolate by themselves in a systematic way.

Category instruments are highly descriptive, replacing observer judgments with data about what happened. Being so specific, they require the observer's close attention and suffer from lapses in attention. If a general overview of class proceedings is important, a SCORE system will not be appropriate, McGreal states.

Observation instruments need not be premade from other sources. Instruments are helpful because they are systematic and relevant for particular uses, not because they anticipate all possible categories of behavior. Observers frequently want to create their own categories to customize their observations while also keeping them focused. Hyman (1986) provides four techniques that can focus any observation. A frequency checklist contains a list of the target behaviors with spaces beside each category to record the number of times each occurred. If an observer targeted questioning behaviors, for instance, one item could be, "Asks the class in general; no student specified," with a space to make a mark when the behavior occurred. Time sampling could be combined with frequency records, showing how many times a particular behavior occurred during a limited period.

A verbatim record, the third technique Hyman describes, keeps track of instances of types of speech. For example, an observer might record teacher questions without classifying or interpreting them. After the data are recorded verbatim, another technique, categorization, provides structure for the postobservation conference and for future observations.

One model of observation--the naturalistic model, often associated with researcher Elliot Eisner (1982)--combines a recognition of objectives with descriptions of the classroom environment. Eisner urges supervisors to structure their observations on two elements: a description of what happens (activities, words, pacing of events, quality of events) and a description of the teacher's characteristic ways of doing things (the teacher's professional style. Observers can take in a more complete picture of the teaching-learning environment by allowing their whole intuitive impressions to take part in the evaluation. Noting only the behaviors of participants without a context can easily mislead observers about how the teacher-student relationship affects learning:

> The average number of soliciting behaviors, the quantitative relationship of teacher talk to student talk, the number of responding to reacting moves simply are not adequate for achieving a conception of how the teacher and the students engage each other. When the characteristics of classroom life are formalized, as

> they are when check-off observation schedules are used, the quality of that life and its meaning for those who are in the situation is radically reduced (Eisner 1982).

Eisner extends this descriptive mode of observing to include the observer's appreciation of the artistry in teaching. He encourages both educational connoisseurship (appreciation of the art of teaching) and educational criticism. Educational connoisseurs have considerable experience in education; they know intimately the thinking and acts of teachers. In their role as educational critics, however, observers aim to lay open the art of teaching--to educate teachers by holding a mirror up to their practice of the art. Educational criticism, in Eisner's view, gives the teacher a vivid image of what the observer saw. The function of the observer, in this approach, is "rendering in artistic language what one has experienced so that it is helpful to the teacher or to others whose views have a bearing on the schools."

Other Sources of Data

Some districts use data besides classroom observations in evaluating teachers. Filling out the perspectives on a teacher's performance can involve parents' and students' evaluations as well as collecting teachers' handouts and assignments. Although the data sources are many, their utility and informativeness may not make them all worthwhile.

Parent Evaluations

Parent evaluations have been probably the least successful. Parents did not respond, for instance, to an invitation from the Berkeley, California, public schools to observe and comment on teachers: only 64 out of a possible 15,000 took up the invitation. Their feedback also contributed little to teachers' knowledge about their teaching (though it may have contributed some knowledge about their students). It would seem, then, that for formative teaching evaluations, parents are not a useful source of information.

Peer Evaluations and Peer Observations

Summative peer evaluations--that is, judgments of teachers' performances by other teachers--have also not proven beneficial. In fact, most research and followup studies indicate that summative peer evaluations are destructive. They harm teacher morale and create lasting grudges among the faculty. Teachers often become testy about peer evaluations: "That's what the administrators get paid for. I'm not going to do their job." "I

refuse to get involved in evaluating people I have to work and interact with everyday." Moreover, it is difficult to find an amenable compromise when teacher and management evaluations clearly differ.

Teachers react positively, however, to peer *supervision*--the formative observation of teachers by their peers. Other terms have been used to describe peer supervision: "colleague consultation," for instance, or "peer consultation." Team teaching includes an element of peer supervision, as teachers share objectives, materials, students, and space.

But other uses of peer supervision occur less "naturally." The structure of schools is not usually conducive to teachers' informal, mutual reviews of their colleagues' work. Thus, peer supervision is being proposed increasingly for special purposes. Articles proposing or reporting peer supervision strategies now appear frequently.

Marginal teachers might benefit from an intensive assistance process developed by Jim Sweeney and Dick Manatt (1984) at Iowa State University. Their proposal involves forming an "intensive assistance team" of faculty members willing and able to coach a colleague. The team performs only formative supervision; the evaluating is left to administrators. The team develops an improvement plan and a target date, recording their work in a log that is also monitored by the principal.

A structured observation plan using faculty can also be used for experienced, competent teachers, though such plans may be most useful for monitoring first-year teachers. The focused team supervision used in Pittsburgh's School Improvement Program concentrates on areas of need, identified through multiple data sources from schools and individual teachers (Bickel and Artz 1984). An approach called "reflective teaching" has teachers teach to their peers and receive feedback on lessons (Cruikshank and Applegate 1981). Carolyn Ruck (1986) has recently discussed the outlines of a collegial supervision arrangement in which the principal acts much like a building contractor--one who coordinates teachers in supervising one another rather than doing all the work himself or herself.

Teaching Materials

Analyzing a teacher's materials can also provide some fertile information for evaluators. Students spend as much time working with teaching materials as they do in receiving instruction from the teacher. In elementary classrooms, students spend 70 percent of their time on such materials, whereas in junior high and high schools the time varies between 40 and 60 percent (McGreal 1983). Thus, improving the delivery of instruction to students should

involve reviewing the effectiveness of teachers' materials, as well as the more usual review of their verbal instruction.

Student Evaluations

Student assessments of teacher performance can be used for evaluations if they are limited to students commenting on the learning climate of the classroom. Teachers are very reluctant to accept students' judgments of their teaching as a valid indicator of success but are often able to credit students with knowledge of the classroom environment.

Gene Glass (1974) has suggested that pupils' evaluations of teachers be one of the three areas of evidence gathered in the evaluation process (the other two being classroom observations and credential information). Such student data should be used to corroborate or contest the observers' ratings of a teacher. They could also inform evaluators about the learning climate in the classroom and "the state of basic human decency that prevails in the classroom," Glass states. As a source of evidence about teachers' performance, student experience could lend authority to other strong data or could raise valid suspicions about a flawed evaluation process.

> Finding that principals' ratings of teacher rapport with pupils do not correlate with pupils' expressions of rapport with the teachers casts doubt on the principals' ratings, the pupils' ratings, or both--and something must be done about the situation. (Glass 1974)

McGreal (1983) notes that teacher rating forms for students are often characterized by general statements about the teacher. Because students' attitudes toward the teacher may fluctuate from day to day, the forms are more likely to record emotional responses than considered thought. The following form is typical of that mistake (this and the following examples are taken from McGreal):

	strongly agree	*agree*	*disagree*	*strongly disagree*
1. The teacher knows the subject matter.				
2. The teacher has favorites.				
3. The teacher is not very interesting.				
4. The teacher emphasizes a lot of memorization.				

A better, more informative questionnaire would focus on students' perception of the learning conditions of the classroom:

1. I feel my ideas are important.
2. Everyone gets a chance to answer questions.
3. I get help when I need it.
4. I am afraid to answer questions.

Self-Evaluations

Like student evaluations, teachers' self-evaluations are best used with caution. Some districts require self-evaluations. They are performed on checklists and then sent to the teacher's file in the central office, gathering dust there. It is an isolated event, without preparation or followup.

Like other sources of data, self-evaluations are most effective when they are a part of a wider array of sources and when they can be discussed with someone else. One use of self-evaluations has teachers compare their own evaluations of their performance with an observer's. This is highly provocative, though, and of dubious value. Certainly, some supervisors may try to anticipate how teachers will rate themselves, in order to prevent having to define a less-complimentary opinion.

Teachers can profit from self-evaluations before the preobservation conference. If a teacher is unsure of what he or she wants to focus on in goal-setting or observation, self-evaluation can point to areas of uncertainty, saving some time in preobservation conferences.

Self-analysis of teaching can also be incorporated into the teacher's ongoing development scheme. A supervisor can help a teacher pick a focus for self-analysis--some aspect of lecturing, discussion, demonstrations, or heuristic approaches. Then, the teacher collects information from tape recordings, videotaping, student feedback, or observations from aides or colleagues. The supervisor and teacher use these self-analysis data in their work together. Acheson and Gall (1987) recommend that a self-analysis goal can be set at the first planning conference of the year and monitored until teacher and supervisor are satisfied.

In one example told by Acheson and Gall, a teacher discovered

that he "put down" students frequently in informal interactions. By recording his informal classroom talk with students several times a week, he noted not only the frequency of his negative remarks but also exactly what he had said. Then, he wrote down alternative phrases he could have used in those situations. His supervisor monitored his progress and agreed to check the perceptions of a few students informally.

Redfern (1980) emphasizes the two-fold nature of evaluation: the teacher evaluating himself or herself, and the observer assessing the teacher. This would be a fair and valid process, though, only if the performance objectives were clear to both teacher and supervisor, and if the expected responses extended only to the behaviors covered by those objectives.

The Dynamics of Feedback

After the lesson has been observed, the teacher and observer may get together to analyze the observational data and set goals for improvement. This is an important occasion in the evaluation process because it allows the teacher to talk in detail about his or her work with someone who has been in the classroom. It is also a time for diplomacy and candor on the part of the observer.

Superficial observations become apparent in the feedback, and unfocused data collection can scuttle efforts at a consensus about what happened in the classroom. It would be wrong, then, to assume that feedback is simply "the tail-end of things." It is a direct outcome of the care with which one takes data and the awareness of teaching the observer brings to the task. At its best, the feedback from observations can be revealing, persuasive, and creative.

Observers can make feedback more useful by eliciting information from the teacher about what transpired in the classroom. An important contribution of clinical supervision models has been to emphasize the teacher's "revealing" role in feedback conferences. Indeed, it would be more accurate to conceive of the feedback going in two directions--being a dialogue rather than a monologue.

The method in the Program for Quality Teaching, for instance, urges observers to listen more and talk less. To do this, the observer can ask for the teacher's feelings, inferences, and opinions, allowing the teacher to do the interpreting. Given data from the observation--a videotape, perhaps, or a written narrative--the teacher asks for the observer's opinion, but the observer turns the interpretation back to the teacher:

T: What do you think of *that*?

Q: Well, what do *you* think of it?
T: I don't like it.
O: OK, then don't do it!
(British Columbia Teachers Federation 1986)

An observer should be sensitive to the opinions hidden behind the teacher's questions, much as a teacher can be trapped by an observer's question. Consider the implications for the teacher if the observer begins a conference with the question "How do you think things went today?" Although the observer appears to be asking an opinion or feeling question, he or she is actually asking for a conclusion from the teacher. The teacher, however, has not yet had an opportunity to examine the data and draw reasonable, informed conclusions; the observer has had that opportunity. To the teacher, then, the situation may seem entrapping, as though the observer were springing a test on him or her.

Clarification can also reveal a teacher's approach and provide valuable information for the observer. How the observer states a question can encourage or discourage a teacher's response and provide varying amounts of information for the observer. The following questions, for instance, seem innocuous enough:

Is Jeannie's behavior different today?
Does noise worry you?
Have they had independent study time before?

The following versions of the questions, though, would probably draw out more information from the teacher and cause no confusion on the teacher's part in answering them:

How does Jeannie's behavior today correspond to her past actions?
Why did that interval of noise seem a problem to you?
When did this group start to study in independent patterns?
(British Columbia Teachers Federation 1986)

The nature of a supervisor's questioning can radically affect the outcomes of a conference. "While two or three well-chosen, well-placed, and well-asked questions may be effective, it is by no means true that twelve or eighteen questions will be six times as effective," states Ronald Hyman (1986). His handbook includes a valuable chapter on question-asking, in which he identifies five types of questions (awareness, information-seeking, delving, divergent, and interpretation/evaluation) and offers some advice on asking questions and fielding teachers' responses. Overall, Hyman's suggestions contribute to give-and-take between teacher and supervisor in an atmosphere of collegiality. To extend one of his assertions to include both asking and listening, the interchanges should take place "in such a way that the quantity and quality of future responses are enhanced," he states.

With that background in mind, observers should first ask their questions in a helpful, positive tone, advises Hyman. Questions meant to raise a teacher's awareness of his or her own behavior may easily seem threatening, laden with implicit judgments. Second, observers should wait for a response after asking a question. They should not answer the question themselves, repeat or rephrase it, or ask another question. Three to five seconds is the minimum time for most listeners to process and acknowledge a good question. To fill in the dead air, observers (like many teachers) are prone to ask a series of questions. (Teachers ask questions on the average of three a minute in the classroom.)

Hyman's third piece of advice is to wait for a response. If the observer wants the teacher to ask questions, the observer should solicit them: "Do you have any questions to clarify the criteria I used in evaluating you?" or "Please ask any questions you need about this new system for gathering data."

Finally, it is important for observers to ask a variety of questions. Hyman notes that observers often have fundamental strategies that they vary in response to the teacher's contributions. Observers may start with awareness questions drawn from the observational data, proceeding then to if-then questions or questions requiring role switching.

Although it may be difficult at times for observers to avoid a leading question, they will be more persuasive by allowing the teacher to draw conclusions or insights directly from the data. The observer's interpretation of the data, without judgment, is often an effective way to emphasize a valid conclusion. In the following example, the observer states a fact and asks a clarifying question, but the teacher herself drew the conclusion:

O: And so this is the pattern of the interaction that took place during the part of the lesson I coded.
T: Do you think I called on Agnes too often?
O: You did call on her more often than anyone else.
T: She seemed to be the best prepared of anyone in class and her answers were good ones.
O: Is this usually the case?
T: No, she seemed unusually willing to respond today. Maybe it was because she was well-prepared or maybe because you were in the room.
O: How did the other kids seem to you to be responding to Agnes?
T: I thought they were agreeing with her and they seemed to be pleased that she was taking the part that she was.
O: Sounds to me like you've answered your own

question.

T: I guess I just wanted you to agree with me.
(British Columbia Teachers Federation 1986)

Because the feedback avoids direct advice to a teacher, the outcomes can be more creative (and thus less perfectly predictable). This method encourages the teacher to suggest alternatives to his or her present mode of teaching. The danger, as the authors of the Program for Quality Teaching note, is that an observer can play superteacher at this juncture, saying, "The logical thing to do now is to repeat that strategy but change the group composition so that. . . ." It is more difficult, but finally more effective, to play a supplementary role: "The observer is better off saying something like, 'What can you think of that might develop the concept more clearly? Let's brainstorm a bit and see what we can come up with.' Thus the observer assumes some risk but does not dominate" (British Columbia Teachers Federation 1986).

It is hoped that eliciting information, being persuasive, and remaining open to alternatives will put the motivation for improvement within the teacher. Observational data are not worth much without that motivation. Ideally, the feedback can contribute to the accuracy of a teacher's self-assessment--an idea appealing to any teacher who wants to become more proficient.

Conclusion

From the energetic discussions found in the literature on teacher evaluation, we can now provide answers for four of the questions asked in the introduction.

First, how can teacher development strategies coexist with accountability strategies? This is probably the major tension now confronting those responsible for or affected by teacher-evaluation programs. It is possible, as we have seen, to accommodate accountability standards and also provide a vigorous development program for all teachers. Such a combination of strategies requires approaching evaluation as essentially a development activity for every teacher and providing special attention to the accountability standards as they affect marginal teachers.

This approach calls for serious, long-term commitment from a school (administrators and teachers) as well as from the central office. It also requires a tactical expertise that accompanies the strategic commitment: that is, trained and competent evaluators are needed who know how to collaborate with teachers in setting individual goals and facing new teaching challenges.

Second, can the same people who decide teachers' career placement also help improve their teaching? If so, how? There are major hurdles for administrators who wish to be both knowledgeable teacher supervisors and objective raters of teachers according to standards. Some researchers have suggested dividing the roles in a school--with supervisors and evaluators (that is, raters of teachers) being different people. There is merit to this suggestion, when it can be done without undue expense.

Evaluators must be distinguished from other administrators or staff by special training and considerable teaching experience. The thorny problems of planning, communication, and feedback that will entangle uninitiated evaluators may destroy their effectiveness and credibility with teachers. One study has found that teachers credit evaluators' lack of training as the foremost barrier to effective evaluations (Stiggins and Bridgeford 1985). Furthermore, the importance of reciprocity in teacher-evaluator relationships must be recognized if an evaluation program is to avoid the hard feelings and futility that so many others have generated.

Third, how useful are evaluation programs for improving teaching? There is no definitive answer to this question. Nor are there any conclusive opinions about which evaluation programs produce the best results. As the model presented in chapter 1 indicates, however, certain features do make evaluation programs

more likely to succeed.

Besides the central track of setting criteria, appraising performance, and communicating the evaluation, other inputs, such as student performance data or teaching artifacts, add valuable information that classroom observation cannot provide. Criteria, too, should take into account research on effective teaching, details of the learning environment of each school and community, and the leadership approaches that each school uses. Finally, two paths emerge from the evaluation process--one is summative (providing the accountability component) and the other is formative (planning between teacher and supervisor for teaching-improvement goals).

From a teacher's standpoint, evaluation is useful if certain conditions are met: (1) attention is paid to teacher input into the process, (2) collegial observation and self-observation are allowed (using videotaping, for instance), (3) feedback is frequent and observations are followed up with goal-setting, (4) the performance criteria are specific and subject to teachers' input when they are formed, and (5) evaluators give detailed suggestions rather than vague criticism or irrelevant observations about teaching.

Finally, what specific approaches to the observations of teachers are the most productive, least time-wasting, and most helpful? For observations to be worthwhile, the experience of practitioners and researchers indicates that they must be carried out by knowledgeable observers using a well-planned, well-recorded set of teaching criteria. Moreover, the results should be diplomatically discussed with the teacher and followed up with further observations and opportunities for teacher self-appraisal.

For the evaluation process to pay back the maximum return on the investment of time and energy, that process should be integrated with goal-setting programs and other developmental activities. When marginal teachers are in jeopardy of being dismissed, the observation/rating process must take place in a context in which the performance criteria have been well known to teachers and ample help has been offered in professional development activities.

Appendix

Training Program for Staff and Supervisors
(Thomas McGreal)

I. Entire Staff (8 hours total)
(Whole-group presentations done by person from outside the district who has worked with the evaluation committee)

A. Introduction to the System (1 hour)

1. Distribute the evaluation pack; the staff sees the system for the first time.
2. Explain the purpose of the system.
3. Present and discuss each part of the system and the requirements for each.

B. Teaching Focus (3-1/2 hours)

1. Provide initial introduction to teaching research.
2. Give examples of the use of teaching research in setting goals.
3. Stress the importance of focusing attention on instruction and on the high level of teacher involvement that the new system encourages.

C. Goal Setting (1-1/2 hours)

1. Discuss the responsibilities of the supervisor and teaching in goal setting.
2. Discuss the approximate time requirements inherent in the new system.
3. Introduce the various types of goals that can be set and how they should be prioritized.
4. Discuss the strategies of goal setting that the supervisors will be taught.
5. Provide a series of sample goals.

D. Data Collection Methods (1-1/2 hours)

1. Discuss the appropriate use of observations and how they will be conducted.

2. Introduce artifact collection and how it is best used.
3. Discuss appropriate uses of student evaluation and include several different samples.
4. Encourage staff to use other alternatives and provide examples of when they might be appropriate (self-evaluation, peer supervision, student performance).
5. Provide sample goals and examples of plans supervisors and teachers might develop to meet goals.

E. Closing Discussion (1/2 hour)

1. Discuss how the system will be monitored the first year....
2. Note that training will be continuous....
3. Ask the staff for their full participation so that the system will have a chance to work.

II. Supervisors (1 day total)

A. Remind supervisors of the importance of attitude to the success of the new system. They must be willing to allow teachers to have equal involvement. They must continually work to display a helping attitude rather than an evaluative one.

B. Review supervisors' specific responsibilities within the system and discuss their approximate time involvements.

C. Specific Skill Training

1. Review goal-setting conference strategies.
2. Practice session: supervisors turn general teacher statements into goals that are focused and manageable.
3. Practice session: supervisors devise appropriate action plans to carry out typical goals.
4. Introduce classroom observation skills.
   a. Supervisors practice their descriptive writing skills.
   b. Introduce and practice using a

series of observation instruments.

5. Introduce conferencing skills.
   a. Review clinical supervision techniques, including suggestions for conducting pre- and post-observation conferences.
   b. Discuss techniques for providing positive and negative feedback.
   c. Supervisors practice writing summative evaluations.

Source: McGreal (1983)

Bibliography

Acheson, Keith A., and Gall, Meredith Damien, *Techniques in the Clinical Supervision of Teachers: Preservice and Inservice Applications.* 2nd ed. New York: Longman, 1987.

Beckman, J. *Legal Aspects of Teacher Evaluation.* Topeka, Kansas: National Organization on Legal Problems of Education, 1981. 70 pages. ED 207 126.

Bickel, William E., and Artz, Nancy J. "Improving Instruction Through Focused Team Supervision." *Educational Leadership*, 41,7 (April 1984): 22-24. EJ 299 429.

Bird, Tom, and Little, Judith Warren. *Instructional Leadership in Eight Secondary Schools.* Final Report. Boulder, Colorado: Center for action Research, Inc., June 1985. 281 pages. ED 263 694.

Blumberg, Arthur. "Supervision in Weakly Normed Systems: The Case of the Schools." Paper presented at a symposium of the Special Interest Group on Instructional Supervision, American Educational Research Association, Montreal, Quebec, April 1983. 19 pages. ED 239 381.

_____. *Supervisors and Teachers: A Private Cold War.* Berkeley, California: McCutchan, 1974.

Bolton, D.L. *Selection and Evaluation of Teachers.* Berkeley, California: McCutchan, 1973.

British Columbia Teachers' Federation. *Program for Quality Teaching: Developing Teaching Practices.* August 1986.

Burke, Peter J., and Fessler, Ralph. "A Collaborative Approach to Supervision." *The Clearing House*, 57, 3 (November 1983): 107-110. EJ 291 221.

Cawelti, Gordon, and Reavis, Charles. "How Well Are We Providing Instructional Improvement Services?" *Educational Leadership*, 38, 3 (December 1980): 236-40. EJ 238 608.

Cogan, Morris L. *Clinical Supervision.* Boston: Houghton Mifflin, 1973.

Cruikshank, Donald R., and Applegate, Jane H. "Reflective Teaching as a Strategy for Teacher Growth." *Educational*

Leadership 38, 7 (April 1981): 553-554. EJ 245 690.

Darling-Hammond, Linda. "A Proposal for Evaluation in the Teaching Profession." *The Elementary School Journal*, 86, 4 (March 1986): 531-551. EJ 337 997.

Darling-Hammond, Linda, and Wise, Arthur E. "Teaching Standards, or Standardized Teaching?" *Educational Leadership, 41,2* (October 1983): 66-69. EJ 286 674.

Darling-Hammond, Linda, and others. "Teacher Evaluation in the Organizational Context: A Review of the Literature." *Review of Educational Research*, 53, 3 (Fall 1983): 285-328. EJ 290 819.

Duffy, Francis M. "Analyzing and Evaluating Supervisory Practice." Paper presented at the annual meeting of the Association for Supervision and Curriculum Development, Chicago, March 23, 1985. 14 pages. ED 254 896.

Duke, David L., and Stiggins, Richard J. "Five Keys to Growth Through Teacher Evaluation." Portland, Oregon: Northwest Regional Educational Laboratory. December 1986.

Eisner, Elliott. "An Artistic Approach to Supervision." In *Supervision of Teaching*, edited by Thomas J. Sergiovanni. 53-66. Alexandria, Virginia: Association for Supervision and Curriculum Development, 1982. ED 213 075.

____. "On the Uses of Educational Connoissuership and Criticism for Evaluating Classroom Life." *The Educational Imagination.* New York: MacMillan, 1979.

Embretson, Gary, and others. "Supervision and Evaluation: Helping Teachers Reach Their Maximum Potential." *NASSP Bulletin*, 68, 469 (February 1984): 26-30. EJ 294 876.

Fitzgerald, James, and Muth, Rodney. "It Takes More than Money to Improve Teaching." *NASSP Bulletin*, 68, 470 (March 1984): 37-40. EJ 294 981.

Garawski, Robert A. "Collaboration Is Key: Successful Teacher Evaluation Not a Myth." *NASSP Bulletin*, 64, 434 (March 1980): 1-7. EJ 217 698.

Garman, Noreen B. "The Clinical Approach to Supervision." In *Supervision of Teaching*, edited by Thomas J. Sergiovanni. 35-52. Alexandria, Virginia: Association for Supervision and Curriculum Development, 1982. ED 213 075.

____. "Reflection, the Heart of Clinical Supervision: A Modern Rationale for Professional Practice." *Journal of*

Curriculum and Supervision, 2, 1 (Fall 1986): 1-24. EJ 341 166.

Glass, Gene V. "Teacher Effectiveness." In *Evaluating Educational Performance: A Sourcebook of Methods, Instruments, and Examples*, edited by Herbert J. Walberg. 11-32. Berkeley, California: McCutchan, 1974.

Haertel, Edward. "The Valid Use of Student Performance Measures for Teacher Evaluation." *Educational Evaluation and Policy Analysis* 8, 1 (1986): 45-60.

Huddle, Gene. "Teacher Evaluation--How Important for Effective Schools? Eight Messages from Research." *NASSP Bulletin*, 69, 479 (March 1985): 58-63. EJ 315 243.

Hunter, Madeline, and Russell, Douglas. "How Can I Plan More Effective Lessons?" *Instructor*, 87 (1977): 74-75, 88.

Hyman, Ronald T. *School Administrator's Faculty Supervision Handbook*. Englewood Cliffs, New Jersey: Prentice-Hall, 1986.

____. *School Administrator's Handbook of Teacher Supervision and Evaluation Methods*. Englewood Cliffs, New Jersey: Prentice-Hall, 1975.

Johnston, Gladys S. and others. "The Relationship between Elementary School Climate and Teachers' Attitudes Toward Evaluation." *Educational and Psychological Measurement*, 5,2 (Spring 1985): 89-112. EJ 320 577.

Joyce, Bruce, and Weil, Marsha, *Models of Teaching*. 3rd ed. Englewood Cliffs, New Jersey: Prentice-Hall, 1986.

Lortie, Dan C. *Schoolteacher: A Sociological Study*. Chicago: University of Chicago Press, 1975. 284 pages.

McGreal, Thomas L. "Effective Teacher Evaluation Systems." *Educational Leadership, 39, 4* (January 1982): 303-05. EJ 257 908.

____. *Successful Teacher Evaluation*. Alexandria, Virginia: Association for Supervision and Curriculum Development, 1983. 175 pages. ED 236 776.

McLaughlin, Milbrey Wallin. "Teacher Evaluation and School Improvement." *Teachers College Record*, 86, 1 (Fall 1984): 193-207. EJ 309 301.

Medley, Donald M; Coker, Homer; and Soar, Robert S., *Measurement-Based Evaluation of Teacher Performance: An Empirical*

Approach, New York: Longman, 1984.

Meyer, John W., and Rowan, Brian. "The Structure of Educational Organizations." In *Environments and Organizations*, edited by Marshall W. Meyer. 78-109. San Francisco: Jossey-Bass, 1978.

Natriello, Gary. *Evaluation Frequency, Teacher Influence, and the Internalization of Evaluation Processes: A Review of Six Studies Using the Theory of Evaluation and Authority. Final Report.* Eugene, Oregon: Center for Educational Policy and Management, University of Oregon, November 1983. 54 pages. ED 242 050.

Peterson, Clara Hamilton. *A Century's Growth in Teacher Evaluation in the United States.* New York: Vantage, 1982.

Peterson, Donovan. "Legal and Ethical Issues of Teacher Evaluation: A Research-Based Approach." *Educational Research Quarterly*, 7,4 (Winter 1983): 6-16. EJ 284 820.

Popham, W. James. *Designing Teacher Evaluation Systems. A Series of Suggestions for Establishing Teacher Assessment Procedures as Required by the Stull Bill (AB 293), 1971 California Legislature.* Los Angeles: Instructional Objectives Exchange, 1971.

Raths, James, and Preskill, Hallie. "Research Synthesis on Summative Evaluation of Teaching." *Educational Leadership*, 39, 4 (January 1982): 310-313. EJ 257 910.

Redfern, George B. *Evaluating Teachers and Administrators: A Performance Objectives Approach.* Boulder, Colorado: Westview, 1980.

Reyes, Donald J. "Bringing Together Teacher Evaluation, Observation, and Improvement of Instruction." *The Clearing House*, 59, 6 (February 1986): 256-58. EJ 331 123.

Reyes, Donald J., and others. "Applying Teacher Effectiveness Research in the Classroom." Northern Illinois University. 1986.

Ross, Doris, and Solomon, Lester. *Evaluating Teachers with Lessons from Georgia's Performance-Based Certification Program.* Denver, Colorado: Education Commission of the States, July 1985.

Ruck, Carolyn L. *Creating a School Context for Collegial Supervision: The Principal's Role as Contractor.* Eugene, Oregon: Oregon School Study Council, Bulletin Series, November 1986, 32 pages.

Sergiovanni, Thomas J., ed. *Supervision of Teaching*. Alexandria, Virginia: Association for Supervision and Curriculum Development, 1982. 207 pages. ED 213 075.

Sergiovanni, Thomas J., and Starratt, Robert J. *Supervision: Human Perspectives* New York: McGraw-Hill, 1983.

Smyth, W. John. "Toward a 'Critical Consciousness' in the Instructional Supervision of Experienced Teachers." *Curriculum Inquiry*, 14, 4 (Winter 1984): 425-36. EJ 310 000.

Stiggins, Richard J. "Teacher Evaluation: Accountability and Growth Systems--Different Purposes." *NASSP Bulletin* (1986): 51-58.

Stiggins, Richard J., and Bridgeford, Nancy J. "Performance Assessment for Teacher Development." *Educational Evaluation and Policy Analysis*, 7, 1 (Spring 1985): 85-97.

Stodolsky, Susan S. "Teacher Evaluation: The Limits of Looking." *Educational Researcher*, 13, 9 (November 1984): 11-18. EJ 309 391.

Sweeney, Jim, and Manatt, Dick. "A Team Approach to Supervising the Marginal Teacher." *Educational Leadership*, 41, 7 (April 1984): 25-27. EJ 299 430.

Wise, Arthur E., and Darling-Hammond, Linda. "Teacher Evaluation and Teacher Professionalism." *Educational Leadership*, 42, 4, (December 1984/January 1985): 28-33. EJ 311 588.

Wise, Arthur E., and others. "Teacher Evaluation: A Study of Effective Practices." *The Elementary School Journal*, 86, 1 (September 1985): 61-121. EJ 324 222.

ERIC®

ERIC Clearinghouse
on Educational Management
College of Education
University of Oregon
1787 Agate Street
Eugene OR 97403

(503) 686-5043

NCREL

North Central Regional
Educational Laboratory
295 Emroy Avenue
Elmhurst IL 60126

(312) 941-7677